万卷方法

混合方法论：定性方法和定量方法的结合

MIXED METHODOLOGY: COMBINING QUALITATIVE AND QUANTITATIVE APPROACHES

阿巴斯·塔沙克里 (Abbas Tashakkori)
查尔斯·特德莱 (Charles Teddlie) 著
唐海华 译
张小劲 校

重庆大学出版社
http://www.cqup.com.cn

万卷方法 是重庆大学出版社自2004年7月以来陆续出版的一套深入、系统地介绍社会科学研究方法的书系，至今已出版图书近60个品种。

“万卷方法”的理想

为社会学、政治学、管理学、经济学、传播学、心理学、教育学等社会科学各领域的学者和研究生提供一个内容规范、使用便捷的“研究方法工具箱”。

“万卷方法”的受众

- 图书馆及大学社会科学各院系的资料室。
- 社会科学各领域的研究人员、研究生及高年级本科生。
- 社会及市场调查的实务工作者。

万卷方法是一个开放的体系，欢迎投稿！

万卷方法与学术规范博客圈：http://q.blog.sina.com.cn/fafang

万卷方法

混合方法论：

定性方法和定量方法的结合

MIXED METHODOLOGY：
COMBINING QUALITATIVE AND QUANTITATIVE APPROACHES

阿巴斯·塔沙克里 (Abbas Tashakkori)
查尔斯·特德莱 (Charles Teddlie) 著

唐海华 译

张小劲 校

重庆大学出版社

Authorized translation from the English language edition, entitled Mixed Methodology: Combining Qualitative and Quantitative Approaches, by Abbas Tashakkori and Charles B. Teddlie, published by Sage Publications, Inc., Copyright © 1998 by Sage Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanial, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. CHINESE SIMPLIFIED language ediiton published by CHONGQING UNIVERSITY PRESS, Copyright © 2008 by Chongqing University Press.

混合方法论:定性方法与定量方法的结合。作者:阿巴斯·塔沙克里,查理斯·特德莱。原书英文版由 Sage 出版公司出版。原书版权属 Sage 出版公司。

本书简体中文版专有出版权由 Sage 出版公司授予重庆大学出版社,未经出版者书面许可,不得以任何形式复制。

版贸核渝字(2006)第 161 号。

图书在版编目(CIP)数据

混合方法论:定性方法和定量方法的结合/(美)塔沙克里(Tashakkori, A.),(美)特德莱(Teddlie, C.)著;唐海华,译.—重庆:重庆大学出版社,2010.3

(万卷方法)

书名原文:Mixed Methodology Combining Qualitative and Quantitative Approaches

ISBN 978-7-5624-5110-5

Ⅰ.混… Ⅱ.①塔…②特…③唐… Ⅲ.①定性决策—分析方法②定量决策—分析方法 Ⅳ.C934

中国版本图书馆 CIP 数据核字(2009)第 163052 号

混合方法论:定性方法和定量方法的结合

阿巴斯·塔沙克里 查尔斯·特德莱 著

唐海华 译 张小劲 校

责任编辑:吴文静 版式设计:雷少波

责任校对:任卓惠 责任印制:赵 晟

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街 174 号重庆大学(A 区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址:http://www.cqup.com.cn

邮箱:fxk@cqup.com.cn (营销中心)

全国新华书店经销

重庆升光电力印务有限公司印刷

*

开本:940×1360 1/32 印张:5.75 字数:164千

2010 年 3 月第 1 版 2010 年 3 月第 1 次印刷

印数:1—4 000

ISBN 978-7-5624-5110-5 定价:23.00 元

本书如有印刷、装订等质量问题,本社负责调换

版权所有,请勿擅自翻印和用本书

制作各类出版物及配套用书,违者必究

作译者简介

阿巴斯·塔沙克里(Abbas Tashakkori)　美国路易斯安那州立大学从事研究方法论教学的教授。他于1979年在北卡罗来纳大学教堂山分校(the University of North Carolina at Chapel Hill)获得社会心理学博士学位,并曾在北卡罗来纳大学教堂山分校的人口学中心从事博士后工作,在德州A&M大学的教育管理学系做过访问学者。他先后任教于伊朗设拉子大学、美国史丹森大学和路易斯安那州立大学的心理学系。

查尔斯·特德莱(Charles Teddlie)　美国路易斯安那州立大学从事研究方法论教学的教授。他于1979年在北卡罗来纳大学教堂山分校取得社会心理学博士学位,此后在新奥尔良大学、英国纽卡斯尔大学和路易斯安那州立大学从事教学工作。此外,他先后在英国、伊朗、比利时、挪威、俄罗斯、乌克兰和白俄罗斯等国就学校效率研究、研究方法论教学等主题进行过学术访问和演讲。

唐海华　2006年毕业于北京大学,获政治学博士学位,现为中国人民大学政治学讲师,对于政治学方法论的教学与科研有相当的知识积累和实践经验。

张小劲　1989年毕业于中国人民大学,获政治学博士学位,现任清华大学政治学系主任、教授,长期从事政治学方法论的教学与科研工作。

前言

本书是对混合方法(mixed methods)和混合模型(mixed model)研究这一发展迅速且备受关注的方法论领域的补苴罅漏之作。尽管当下已有若干位作者正在多个领域内就混合方法问题有所撰述,但是他们的许多文章和章节仍缺乏明晰的概念,他们也往往没有对正在应用的方法举证出具体的案例。就此而论,本书的独特贡献在于两个方面:第一,对混合模型和混合方法研究,我们构建了一个在逻辑上周全详尽的分类体系;第二,针对每一种主要的研究类型,我们提供了可用于实践的操作指南。

尽管作者开始在上述多个领域就这些新的方法发表了相关文字,但定量传统的主流学者(如实证主义者和后实证主义者)以及定性传统的主流学者(如建构主义者),都还没有接受混合模型的研究。譬如,著名的后实证主义的"圣经"式著作(Cook and Campbell,1979)仅仅是在讨论单一方法的偏畸时才提及了混合方法。与此相似,《定性研究手册》(Denzin and Lincoln,1994)直接论及混合方法问题的只有一章的篇幅(即关于临床诊断研究方法的一章)。

混合方法的主要研究文献目前在多个领域中已经得到了长足的发展,但很少有作者尝试过对这些方法作出类型学分析(参见Creswell,1995;Miller and Crabtree,1994;Morse,1991)。社会科学和行为科学的一些主要领域和次领域已经要求采用混合方法。这些领域包括教育、评估、护理、公共健康、社会学、临床研究、管理科学、社区心理学、女性研究和学校效率研究等。

这些不同领域对混合方法和混合模型研究的应用存在着相似性。在本书的第一部分,我们对混合方法和混合模型研究进行了逻辑分

类，指出了我们在这些不同领域中所辨识出的混合方法之间的相似性。我们明确区分了混合方法与混合模型：混合方法是指在一项研究的方法论中（如在资料收集阶段）融合了定性研究和定量研究的路径，而混合模型研究则是指在研究过程的所有阶段（如概念化、资料收集、资料分析和推论）中都融合了定性和定量两种路径。在对相关文献进行了精细的分析之后，我们认为，混合模型研究是社会科学和行为科学研究中日益强化的一种发展趋势。

在本书的第二部分，我们列出了混合方法和混合模型研究的各种方法和策略（譬如抽样、测量、资料收集、资料分析）。其中，我们首先按照这些主题概括归纳了传统的研究路径，然后我们展示了研究者们在研究过程中如何将不同的研究技术融合在一起。在这一部分，我们列举了从多个领域中挑选出的混合方法论路径的特别案例。在挑选这些案例时，我们并未对研究的整体质量或水平作出判断，它们之所以被展列在这里只是为了说明多个领域正在应用着的这种方法。此外，我们可以从多个学科中找到成百上千的应归类为混合型的研究案例，但此处只能展列出一小部分。

在本书第三部分，我们进一步列举了关于混合模型设计的简单案例和扩展案例。为了提高本书对研究生或混合研究的首次使用者的参考价值，这些案例中收入了几篇我们自己的学生近期所撰写的论文。这些论文大多采用了顺序型混合方法（sequential mixed methods）或平行型混合方法（parallel mixed methods），即交替使用或同时使用定性和定量两种基本路径去分析同一个现象。这些案例进一步精确地说明了这些论文的研究及其他的研究是如何进行混合设计的。

我们认为，混合方法和混合模型研究尚处于初级阶段，其发展受到了本书在第一部分所反思的范式之争后遗症的延阻。如今，这一争论已经离我们远去，研究者开始自由地使用最适合于他们所研究问题的各种方法。同时，我们还认为进行研究时应该确立明晰的目标，无论是回答一个疑问、解决一个问题或是评估一个项目；我们强调，相对于范式而言，研究的问题才更为重要，更具有优先性；我们主张，研究者为了回答其研究的问题，应当同时从定量和定性两种路径中选择其合适的方法。就社会科学和行为科学的大多数课题而论，其研究的问题大都可以用混合方法或混合模型设计来求得最佳的解答，远胜于单一地依赖于定性研究或定量研究路径。

致谢

我们对我们的家人、同事以及学生抱有感激之情，感谢他们在本书写作期间所赐予的支持和理解。我们还必须感谢 Marylyn Sines 为本书所提供的编辑协助；感谢 Spencer Maxcy 教授针对第一部分所提供的建议和评价，感谢 Mindy Crain-Dorough 所提供的图书查询。同时，我们还要感谢 Sharareh Nikbakht 所准备的图示和图形。

目录

第一部分　研究中的范式与政治

1　**导论:社会和行为科学中的混合方法与混合模型研究** …… 2
范式之争与混合方法论 …… 2
关于范式之争的更多详情 …… 5
方法论路径在社会和行为科学领域内的演变 …… 12

2　**实用主义与研究策略的选择** …… 19
研究问题主宰,而非范式或方法为王 …… 19
范式之比较 …… 21
与单一方法有关的设计问题 …… 29
社会和行为科学领域中传统资料收集方法的类型学 …… 35

3　**混合方法和混合模型研究的研究设计问题** …… 38
与混合方法研究相关的设计问题 …… 39
与混合模型研究相关的设计问题 …… 49

第二部分　研究的方法与策略

4　**抽样、测量及推论质量** …… 59
抽样及其与通则化/推论之质量的关系 …… 62
测量的质量和推论的质量 …… 75
信息/资料的质量:定量研究的视角 …… 76

推论/结论的质量:定量研究的视角 …… 83
资料的质量与推论:定性研究的视角 …… 88

5 资料收集策略和研究程式 …… 92
资料收集的传统路径和替代路径 …… 92
资料收集策略 …… 97

6 传统资料分析策略的替代性选择 …… 109
传统的定量资料分析 …… 109
传统的定性资料分析 …… 114
混合方法研究的替代性分析策略 …… 122

第三部分 混合模型研究的运用、案例和未来的方向

7 混合模型设计的案例 …… 134
确证性研究 …… 136
探索性研究 …… 140

8 混合模型设计的扩展案例 …… 144
一个平行型混合模型研究(类型Ⅶ)的案例 …… 148
一个顺序型混合模型研究的案例 …… 149
在单项研究中分步骤进行简单和复杂的混合模型设计的案例 …… 151

第9章 结论与展望 …… 162
参考文献 …… 165

译后记 …… 171

研究中的范式与政治

第一部分

这一部分将讨论社会和行为科学研究中的范式与政治。第1章和第2章将回顾范式之争的主要问题。范式之争即围绕着相互对立的世界观和信仰体系就彼此的优缺点所展开的论争，而在社会和行为科学领域，这种争辩主要发生在实证主义者和建构主义者之间。同时，在第1章和第2章，我们还将讨论另一种范式——实用主义。在第2章，我们将进一步区分四种主要的科学范式，即逻辑实证主义、后实证主义、实用主义与建构主义。

在本部分中，将对社会和行为科学的方法论路径进行分类：包括单一方法、传统意义上的混合方法，以及一种趋于将定性研究路径和定量研究路径结合起来的独特方法——我们称之为混合模型研究。混合方法是在研究的具体方法上将定性研究路径和定量研究路径结合在一起，而混合模型研究则是在研究过程的所有阶段中都将这两种路径结合起来。

立基于三个主要研究步骤(即确定研究问题/假设、资料收集和研究操作、分析和推论)的差异性，第3章对混合模型研究进行了新的分类，这又成为本书其余部分篇章安排的依据。

1

导论：社会和行为科学中的混合方法与混合模型研究

范式之争与混合方法论

范式之争举隅

在过去30年里，就两种主要范式或模型的高下优劣问题，社会和行为科学领域中业已发生过多次辩论或“争战”（如 Datta, 1994; Gage, 1989; Guba and Lincoln, 1994; House, 1994; Rossi, 1994）。这两种模型，一个是众所周知的实证主义/经验主义路径，另一个则是人们所熟知的建构主义/现象学路径（如 Cherryholmes, 1992; Guba and Lincoln, 1994）。

所谓“范式”，可以界定为引导研究者的世界观或信仰体系（Guba and Lincoln, 1994）。范式在今天的社会和行为科学中所被赋予的重要地位起源于库恩在1970年发表的名著——《科学革命的结构》。他在这本书中说明，范式是在任一特定领域中都能得到重复使用的模型，相互竞争的不同范式可以同时存在，尤其是同时存在于并不成熟的科学学科之中（Kneller, 1984; Kuhn, 1970）。

实证主义范式强调所谓的定量研究方法，建构主义范式则主张定性研究方法（如 Guba and Lincoln, 1994; Howe, 1988; Guba and Lincoln, 1985）。因此，关于这两种范式的争论有时亦被称为定性—定量之争（如 Reichardt and Rallis, 1994）。缩写词 QUANs 与

QUALs 用来指代这些争辩或“论战”的参与者，QUANs（定量论者）是指那些倾向于定量观点的人，QUALs（定性论者）则指倾向于定性观点的人。

这种范式之争的战火波及多个重要的概念性问题的“战场”，譬如“现实的本质（nature of reality）”、“因果联系的可能性（possibility of casual linkages）”等。社会和行为科学领域中没有哪个学科可以在这场范式之争中袖手旁观而不必表明立场。Datta（1994）将这场战争中的参与者称为“摔跤手（wrestlers）”，而我们更愿使用另一个词——“斗士（warriors）”。

教育学中的“斗士”有 Guba 和 Lincoln（1985），他们认为实证主义的原则以及与之相关联的定量方法论已经不再可信。这两位作者声称，建构主义和定性方法正冉冉上升。Smith 和 Heshusius 也在教育学领域著书立说，他们在 1986 年主张对两大阵营的沟通“关上大门”，因为这两种范式的不相容性使得继续对话毫无意义。因而他们的观点又被称为“不相容论”（incompatibility thesis）。

类似的，心理学在 1970—1980 年代也见证了学者 Cronbach（1982）、Cook 和 Campbell（1979）之间所发生的大规模方法论之争。这场争论主要集中于内部效度（强调实证主义者视之如神圣的受控环境（controlled settings））与外部效度（强调建构主义者趋之若鹜的自然环境（natural settings））之重要性的高下优劣问题。1973 年，Gergen 也质问道：“社会心理的科学性是可能的吗？”他对支撑着心理学各个分支领域的实证主义教义，特别是超越时间—场景之通则的可能性问题提出了强烈的质疑。

而在人类学领域，Gardner 在 1993 年对 Margaret Mead 在萨摩亚群岛的人类学研究工作进行了指责，理由是 Mead 过度依赖于预设性的概念，而且还天真地接受所谓关键受访者的信息。这些反应典型地体现了实证主义阵营对建构主义或“自然主义者”阵营的批判。

关于这场争论的最后一个范例则发生在评估研究的应用领域。Datta 于 1994 年指出，美国评估研究协会连续三任主席（Fetterman，1992；Lincoln，1991；Sechrest，1991）的“对话”，在维护自己的方法论立场并攻击与其相对的“另一派”上，都显得非常强

硬。尽管这样的争辩在所难免,然而在 1980 年代和 1990 年代早期,却日益丧失其实际效果。

范式之争的结束与混合方法的出现

针对社会和行为科学领域的两大范式之争,许多人也曾尝试提出和解之道。"和平主义者"由此开始形成,他们认为定性研究方法和定量研究方法实际上是可以相容的。在教育学和评估研究领域,一些作者(如 Howe,1988;Reichardt and Rallis,1994)立基于一个不同的范式,提出了相容理论,也有人将这一独特的范式称之为"实用主义"。就此而论,对于范式之争中的和平主义者,我们也可以称之为"实用主义者"。

正是在这一时期,范式之争在社会科学哲学的历史上第一次开始变得相对中庸了,因为许多活跃的理论家和研究者都开始采取"范式相对主义"的立场,或对研究中的特定问题随意采用某一种哲学和方法论路径(如 Howe,1988;Reichardt and Rallis,1994),甚至连一些最著名的斗士(即 Guba 和 Lincoln)也发出了息事宁人、结束争战的信号。他们声明:

> Gage 于 1989 年所形容的范式战争的隐喻,毫无疑问是过头了。将过去 10 年、20 年中的讨论和口角形容为战争,已经使事情被涂抹成远远超出必要程度的对抗。比现有任何一种范式更加圆通和精巧的新范式出现之际,便是范式间分歧的解决之时。而这种情形只有在这几种不同观点的持有者聚在一起讨论他们之间的分歧时才最有可能发生。(Guba and Lincoln,1994:116)

如今,实用取向的理论家和研究者们提出了"混合方法"(亦可称"混合方法论"或"方法论混合"),力图同时包括定性路径和定量路径的因素(如 Brewer and Hunter,1989;Patton,1990)。例如,Greene,Caracelli 和 Graham 在 1989 年就展示了运用混合方法所完成的 57 项研究,并且描述了这些混合研究的设计特性。我们将在本章稍后进一步讨论混合方法的具体类型问题。

当前的状态

我们接受范式相对论中所暗含的假定,即范式之争已经结

束，并且被上面简要提到的实用主义取向所替代。正如 Brewer 和 Hunter(1989:22)所认识到的，大多数社会和行为科学的主要领域如今都将多元方法的使用视为理所当然的事情："1950 年代以来，社会科学得到了极大发展。伴随着这些发展，如今实质上已经没有任何重大题域是死守在一种方法之内展开研究工作的。"

范式之争的缓解对许多领域的研究发展具有积极的意义，因为大多数研究者如今都主张不论哪种方法，只要适合研究就行，而不再死抱着某一种方法。然而，实用主义者在描述他们的方法论时，却经常使用含混的语言。他们会使用一些相当宽泛的词汇(如混合方法)来表示在进行一项研究或一系列研究时所遵循的若干种不同的方式。Datta 在提到这种她所谓的"杂糅模型"(mixed-up models)时指出，这源于"世界观的回避与实用主义或混合模型研究的理论"。她的结论是，"这种理论尚有待于得到透彻的阐释"(1994:59)。

我们不敢自称能在这本小册子中为混合方法和混合模型研究提出一个正式的理论，但我们的确希望能够为将这种研究更系统地概念化提供一些指导方针。在介绍我们对混合方法和混合模型研究的分类之前，我们在本章接下来的部分将简单介绍与范式之争及其解决相关联的一些主要问题。

关于范式之争的更多详情

以下所述涉及范式之争及其如何获得使社会科学共同体多数成员感到满意的解决方法，这是专为那些对此尚不熟悉的读者所作。对范式之争历史的简要回顾并不是要在科学哲学(philosophy of science)上展开鸿篇大论，而仅仅是对范式之争及其结局作一番走马观花——对这些问题已经熟悉的读者当然可以跳过这一部分。

有关范式的争论在历史上具有重要的意义，这从其持续时间之长便可略见一斑。Hammersley(1992)注意到，关于定性研究和定量研究的争论实际上起源于 19 世纪中期；20 世纪 20 年代和 30 年代在社会学中浮现；而最近这一争论重新获得关注始于 20 世纪 60 年代定性研究偶尔在社会学和心理学中的复兴，因为这两个领

域在整个20世纪40年代和50年代都被定量研究(调查或实验)所支配(Hammersley,1992)。

尽管存在着这两种主要的对立主张,但显然,人们所提出和辩护的哲学取向或哲学范式远不限于此(如Greene,1994;Guba,1990;Guba and Lincoln,1994)。我们将提到的有四种哲学取向,即逻辑实证主义、后实证主义、实用主义和建构主义(还有其他一些变体,如解释主义、自然主义等)。我们之所以选择这些取向展开讨论,是由于它们代表了范式之争各个主要阶段的不同方面:①第二次世界战后对逻辑实证主义的批判;②后实证主义的盛行;③建构主义的崛起,之后即发生范式之争;④实用主义和相容理论。本章接下来的四个小节将简要地描述其中的每个阶段。

第二次世界大战后对逻辑实证主义的批判

实证主义(也称逻辑实证主义)的源头可以追溯到19世纪的法国哲学家August Comte。实证主义将知识仅仅建立在可观察的事实上,拒绝对“终极起源”的猜想。Lincoln和Guba为实证主义归纳了这样几条公理:

1. 本体论(现实的本质):实证主义者相信存在着唯一的现实。

2. 认识论(认识主体与认识客体的关系):实证主义者认为认识主体与认识客体之间是相互独立的。

3. 价值论(价值在研究中的作用):实证主义者认为研究是价值中立的。

4. 通则化:实证主义者认为超越时间和情境的通则化是可能的。

5. 因果联系:实证主义者认为在结果之前或同时,必有真正的原因。

此外,我们还要增加一条为许多作者(如Goetz and LeCompte,1984;Patton,1990)所提到的第六条公理:

6. 演绎逻辑:强调从一般到特殊的推论,或强调预先的假说(或理论)。

第二次世界大战以后,逻辑实证主义作为一种科学哲学已不

受信任(如 Howe,1998;Phillips,1990;Richardt and Rallis,1994)。1950 年代和 1960 年代,整个社会和行为科学领域中对实证主义的公理(尤其是关于本体论、认识论和价值论的)的不满变得越来越流行。正如 Guba 和 Lincoln(1994)所言,后实证主义是实证主义的知识后裔,并从这时开始致力于解决那些备受诟病的实证主义原则。

后实证主义的盛行

后实证主义的标志性著作(如 Hanson,1958;Popper,1959)出现于 1950 年代末期,很快就在整个社会科学共同体中赢得了广泛的声誉。后实证主义对广受质疑的实证主义公理作出了反应,它的许多立场与其实证主义前身正好相反。

尽管在 1950 年代和 1960 年代许多定量研究者仍然采用实证主义的原则,但是 Reicharde 和 Rallis 令人信服地指出,当时一些最有影响力的定量方法论者在研究取向上却是“不加掩饰的后实证主义者”。Reicharde 和 Rallis 认为,这些方法论者是后实证主义者,因为他们的著作表明他们赞同以下后实证主义的原则:

- 探索中渗透着价值(value-ladenness):研究受到研究者价值观的影响。
- 事实中渗透着理论(theory-ladenness):研究受到研究者使用的理论、假设或框架的影响。
- 现实的本质:我们对现实的理解是建构的。(关于这些问题的进一步讨论见第 2 章)

现在,这些后实证主义者的原则已被定量取向的研究者和定性取向的研究者所共享,因为它们更好地反映了人们在 20 世纪后半期对“现实的本质”以及社会行为研究操作的共同理解。Reicharde 和 Rallis 断定,相较于逻辑实证主义,后实证主义更准确地描绘了当代定量研究的特点。他们指出,一个定量取向的后实证主义阵营已经形成,其中包括了 1960 年代和 1970 年代一些最著名的定量取向研究者。

例如,实验心理学者 Rosenthal(1976)花了很长篇幅讨论什么是“实验者效应”:实验者的观察、感觉或者行为,都可能会在无意

中影响研究的结果。Cook 和 Campbell(1979)是两位确定无疑的定量取向研究者,他们承认实验者的偏见会损害研究的效度:这种实验者偏好充分地展现了方法论方面所存在的缺陷,而这一缺陷会导致事实中充满了价值观念或理论(这正是后实证主义者的基本观点)。Cook 和 Campbell(1979:24)在讨论社会科学的因果联系时指出:

> 我们同意后实证主义的观点:观察是受到理论渗透的,为表达数据而建构精细的科学工具和程序时常常会或明或暗地接受一些很成熟的科学理论,而这些理论超越或涵盖着正在被检验的理论。

与此相似,实验社会心理学家们(如 Jones and Davis,1996;Jones and Nisbett,1972)也研究和讨论了个性因素(dispositional attributions),这涉及行为者—观察者在感知上的不同。就人格理论(personality theory)的检测是否会在操作中受到研究者个人因素(如价值观)的强烈影响,他们进行了深入的思考。这样一个案例再次说明了那些使用传统实验方法和定量数据的研究者也在积极地探索和讨论后实证主义的原则,如价值渗透事实等。

由此,在 1960—1980 年,定量方法论者对一些后实证主义的原则进行了论述,并提供了相应的经验证据。但是,尽管这些方法论者支持那些与后实证主义相关的假设,但不要忘记,他们是在一个强调"方法论正确性"的传统内从事研究工作的(Smith,1994)。因而当面临着在定性取向和定量取向之间择一而行时,这些方法论者倒向了实证主义,对堪称实证主义之标志的实验设计表示支持。例如,Cook 和 Campbell 强硬地回应了对其研究效度特性的批评意见,断言"我们认为,读者们相信因果推论是重要的;而在获知原因的多种方式中,实验方法即使不是最有用的,也是最有用的之一"(Cook and Campbell,1979:91,原文是斜体)。

在这本出版于 1979 年的名著中,显然可以读出两位作者对实验型(或准实验型)研究及其价值内部效度的高度青睐,以及他们对因果推论在严格限定的背景中才可能展开所持有的信念。

建构主义在范式之争前的崛起

实证主义的声誉江河日下,这又导致了比后实证主义更为

“激进”的范式日益流行。这些范式名称不一，诸如建构主义、解释主义、自然主义等，不一而足，但建构主义是最为流行的一种。与这些范式相关的理论家们借鉴了后实证主义，但又向模型中注入了他们自己的维度（如 Denzin, 1992; Gergen, 1985; Goodman, 1984; Hammersley, 1989; LeCompte and Preissle, 1993; Schwandt, 1994）。一些建构主义理论家不满于看到实证主义、后实证主义与他们自己的哲学取向平起平坐，认为必须着力证明他们自己的范式才是最优越的，用以克服实证主义和后实证主义根深蒂固的传统中所包含的偏见。例如，Lincoln 和 Guba（1985:32）对一位著名的定量方法论者试图调和这些不同观点的努力提出指责：

> 一些学者认为后实证主义不过是一种过度反应罢了，如今应当将实证主义与相对主义（后实证主义的特点）修好合一了。Donald T. Cambell 就是其中一位。他提出，进入一个后-后实证主义的时代正当其时，实证主义与后实证主义将结秦晋之好，永世不易。

Lincoln 和 Guba 遂列举了实证主义范式与自然主义（他们所理解的建构主义）范式之间的诸种不和，意在使二者的“秦晋之好”变成镜中花、水中月。参照前述实证主义的五条公理，他们设立了如下五项自然主义范式的原则：

1. 本体论（现实的本质）：自然主义者相信现实存在是多元的、建构的。
2. 认识论（认识主体与认识客体之间的关系）：自然主义者认为认识主体与认识客体是不可分割的。
3. 价值论（价值在研究中的作用）：自然主义者认为研究是受价值制约的。
4. 通则化：自然主义者认为超越时间和情境的通则化是不可能的。
5. 因果联系：自然主义者认为区分原因与结果是不可能的。

一如前文那样，我们在此还须补充为许多作者所提到的第六条特性：

6. 归纳逻辑：强调从特殊到一般的推理，或强调“扎根”理论

("grounded" theory)。

二者若如此黑白分明,那么,在 Smith(1994)所谓持有"范式纯粹(paradigm purity)"之立场的个体之间爆发范式之争也就是难免的了。举例说,Guba(1987:31)声称,上述两种范式非此即彼,正如"相信世界是圆的,也就是不相信世界是平的"一样。Lincoln 和 Guba(1900,1994)不止一次地强调两种范式之间在本体论、认识论和价值论上的龃龉,这无疑是在为范式之争加油添火。Smith(1983:12)曾这样阐述(方法论的)不相容论:

> 其中一种路径在主观问题上采取了主体—客体立场,另一种路径则采取了主体—主体立场;一种路径将事实与价值分开,另一种路径则认为它们之间是难解难分的混合体;一种路径意在探索规律,另一种路径则只为寻求理解。两种立场看上去格格不入,并不相容。

这些范式"清教徒"进一步在其研究方法中论说其不相容论:由于作为研究方法之基石的范式之间互不相容,因而定性方法与定量方法之间的融合也是不可能的。根据这些理论家的观点,那些试图将两种方法融合起来的研究者注定要失败,因为二者所凭依的哲学具有内在的分歧。

实用主义与相容理论

许多有影响的研究者已经指出,这两种范式之间的分歧被强调过头了,二者之间的裂痕并不像"纯粹主义者"所描绘的那样巨大深刻。例如,House(1994)在指出实证主义传统与建构主义传统各自的优缺点时断定,这种二分法是来自于"对科学的误解"。他进一步指出,"通向目标彼岸的方法论道路并不是既定的"(House,1994:20-21)。

有许多言之有理的缘由可以昭示范式之争的缓和。例如,在就评估问题撰文时,Datta(1994)提出了五个令人信服的实践性原因来证明这两种方法论及其背后的范式之间可以"共处"下去:

- 事实上,这两种范式都得到了长期的使用。
- 许多评估者和研究者都强烈要求同时使用这两种范式。
- 研究资助机构同时对这两种范式表示支持。

- 这两种范式都会影响到政策。
- 这两种范式都使我们受益匪浅。

在哲学层面上，实用主义者不得不与范式斗士的不相容理论筑垒对抗，这一理论所依据的是认识论与研究方法之间的关联性。为了对付这种范式—方法的关联论，Howe(1988)选择了使用另一种不同的范式：实用主义。Cherryholmes(1992)和 Murphy(1990)已经将实用主义的起源追溯到 C. S. Peirce、William James 和 John Dewey 等美国学者，以及其他一些更加晚近的理论家，如 W. V. Quine、Richard Rorty 和 Donald Davidson。但这一哲学几乎只在其起源地(美国)得到认同，欧洲学者总有点蔑视实用主义，因为它对形而上的概念(如真理)，持有批判的态度(如 Nielsen, 1991；Rorty, 1990)；与寻求形而上的真理相反，实用主义者将真理看成是"起作用的东西"。Howe(1988:15)将这种取向总结如下：

> 毕竟，大多数实用主义哲学(如 Davidson, 1973；Rorty, 1982；Wittgenstein, 1958)都是解构性的——试图劝说哲学家们不要使用诸如"真理"、"现实"和"概念图景"之类的概念，不再将这些概念转变成"真理"、"现实"和"概念图景"之类的宏大概念并在这一过程中提出一些无法解决的伪问题。(原文是斜体)

Howe 所理解的实用主义的一个主要原则就是，定量方法和定性方法是相容的。这样一来，由于实用主义宣称这两类方法是相容的，因而研究者就可以在他们的研究中同时使用这两种方法。Brewer 和 Hunter(1989:74)的观点与此如出一辙：

> 但是，通过提出不同的具体问题，实用主义在对同一个总问题的研究中可以使用多种研究方法，这对于社会理论具有一些实用的意义。与其固守一种特定的理论风格……及其最相容的方法，不如将各种方法结合起来使用，而这会鼓励，甚至要求将不同的理论视角整合起来用以解释研究资料。

在分析"定量研究与和定性研究"的相容性时，Reichardt 和 Rallis(1994)走得更远。他们指出，定性研究者与定量研究者的基本价值观之间有大量的相似性，足以"结成持久的合作关系"(Reichardt and Rallis, 1994:85)。这种基本价值观的相似之处包

括相信价值渗透研究、相信理论渗透事实、相信现实是多样的和建构的、相信知识是会有错的,以及相信事实无法为理论提供充分的证明,等等。

前三个相似性在本章的前面部分已经有所讨论。而Reichardt和Rallis(1994)认为,定量研究者(与定性研究者一样)也相信知识是会有错误的,并引用了Cook和Campbell(1979)的话,“我们无法证明某一理论或其他因果命题”(Reichardt and Rallis,1994:22)。两位作者进一步指出,定量研究者与定性研究者同样都认为事实无法为理论提供充分的证明,换言之,“任何一套既定的研究资料,都可以用多个理论来解释”(1994:88)。Reichardt和Rallis还同时列出了定量研究者与定性研究者在评估领域中的其他一些共享观念,它们关涉到理解和改善人类状况的重要性,交流结果对于通达明智之决定的重要性,以及相信“世界是复杂的、深奥的,常常难以理解”(1994:89)。

由此可见,许多社会和行为科学家之间存在着一套共同的信念,这些共同信念从根基上催生了有别于实证主义、后实证主义和建构主义的又一种范式,即所谓的实用主义。正是这种实用主义范式使研究者得以在社会和行为研究领域运用混合方法。

方法论路径在社会和行为科学领域内的演变

正是这样,许多社会和行为科学家,逐渐在实用主义哲学取向背后的基本假设问题上形成了共识。在这些基本的哲学问题已经得到讨论的基础上,我们现在可以将注意力转到与混合方法和混合模型方法相关联的、具体的方法论问题之上。

来自不同学科的各类学者(如Blalock,1978;Brewer and Hunter,1989;Datta,1994;Patton,1990)已经在呼吁,我们目前在社会和行为科学领域所使用的不同的方法论在描述方面应当更加协调统一。Brewer和Hunter(1989)也非常强调个体研究者(和研究组)在研究中综合使用不同方法的必要性,并特别呼吁要建立一个更具综合性的方法论路径。

这种方法论的整合努力的其中之一关涉到要使我们用以描述多样方法的语言变得更为精确。对于在社会和行为科学中所

使用的方法，研究新手面对的确实是一系列令人迷惑的名称：单一方法（即定量方法和定性方法，以及这两种方法各自所有的变体）、多样方法（multiple method）、混合方法、多元方法研究（multimethod research）、三角测量法（triangulation of method）、方法论混合等。因此，我们将在本节对方法论路径提出一个分类法，并简单回顾这些方法论路径的演进过程。

表1.1就是对社会和行为科学方法论路径的分类法。首先是三个大类的划分：单一方法（从19世纪社会科学的出现起，一直到1950年代）、混合方法（产生于1960年代，至1980年代变得更加普遍）和混合模型研究（1990年代作为独立的类型出现，但此前已有一定的发展）。

表1.1 社会和行为科学领域方法论路径的演变

时期Ⅰ：单一方法或“纯粹主义者”时代（大约从19世纪到1950年代）

A. 纯粹定量取向
- 1. 单一资料来源（定量研究）
- 2. 在一个范式/模型内，多个资料来源
  - a. 顺序的（定量研究/定量研究）
  - b. 平行的/共时的（定量研究+定量研究）

B. 纯粹定性取向
- 1. 单一资料来源（定性研究）
- 2. 在一个范式/方法内，多个资料来源
  - a. 顺序的（定性研究/定性研究）
  - b. 平行的/共识的（定性研究+定性研究）

时期Ⅱ：混合方法的出现（大约从1960年代到1980年代）

A. 同等地位设计（两种范式/方法都使用）
- 1. 顺序的（即两阶段的顺序研究）
  - a. 定性研究/定量研究
  - b. 定量研究/定性研究
- 2. 平行的/共时的
  - a. 定性研究+定量研究
  - b. 定量研究+定性研究

B. 主次设计（两种范式/方法都使用）

续表

1. 顺序的
a. 定性研究为主/定量研究为次
b. 定量研究为主/定性研究为次
2. 平行的/共时的
a. 定性研究为主 + 定量研究为次
b. 定量研究为主 + 定性研究为次
C. 多层次路径设计
时期Ⅲ:混合模型研究的出现(大约在1990年代)
A. 在研究阶段中单一使用*
1. 研究的类型——定性研究或定量研究
2. 资料收集/操作——定性研究或定量研究
3. 分析/推论——定性研究或定量研究
B. 在研究阶段中多元并用**
1. 研究的类型——定性研究并/或定量研究
2. 资料收集/操作——定性研究并/或定量研究
3. 分析/推论——定性研究并/或定量研究

* 必须每种方法至少在研究的一个阶段中出现。

** 必须两种方法至少在研究的一个阶段中同时出现。

社会和行为科学领域方法论路径演变的主要阶段

我们将在本节说明,社会和行为科学领域中存在着一个从使用单一方法到使用我们所说的混合模型研究的演变过程。而对这一演变过程的讨论中将着重于三个要点:①这一演变过程首先涉及的是对混合方法之应用的接受;②这一演变过程涉及在范式之争中所出现的区分运用于研究进程的所有阶段;③这一演变在过去30年里不断以加速度的方式推进。

演变的第一阶段:从单一方法到混合方法

这一阶段涵盖了从仅仅使用某一种基本的科学方法转变到同时使用多种方法。因此,在表1.1中,这体现为从时期Ⅰ过渡到时期Ⅱ。范式之争以及经由实用主义与相容理论所实现的终结,

导致了这一转变的出现。关于这一过程的历史，本章此前的部分已经有过讨论，在第 2 章和第 3 章，我们还将对纯粹单一方法与混合方法的不同类型进行更为完整的描述。

演变的第二阶段：从混合方法到混合模型研究

这一阶段涉及从仅仅在方法上考虑区分问题，发展到在整个研究过程的所有步骤都考虑区分问题。如表 1.1 所示，这是指从时期Ⅱ到时期Ⅲ。尽管多元方法的出现通常被看成只是一个方法论问题，但范式之争当中所形成的认识论与方法的关联性意见使得有必要考虑不同的取向如何影响着研究过程的其他步骤（如问题形成、研究设计、资料分析和资料阐释）。一如 Howe（1988：15）所言，“定量—定性的区分被运用于多个层次：资料、设计和分析、结果的阐释和认识论范式”。

Creswell（1995：176）则提出了一个关于将范式—方法的关联性运用于研究过程其他阶段的基本问题：

> 要最有效地同时运用这两种范式，就意味着要走向下一步的融合设计：在单一的研究中，除了方法（如对研究的导入、文献与理论、目标陈述以及研究问题）之外，设计过程的各个方面也能够从不同的范式中推导出来吗？

他的回答是肯定的。同时，他还提供了一些例证，证明不同的范式或观点是能够运用于研究过程的这些阶段。

与此类似，Brewer 和 Hunter（1989）将他们的多元方法路径运用到了研究过程的所有阶段，而不仅仅是测量阶段。他们所讨论的阶段包括问题的形成、理论的建立和检测、抽样、资料收集和分析以及报告。他们总结道：

> 采用多元方法路径来进行测量的决定，不仅影响到测量本身，而且影响到研究的所有阶段。确实，通常只在研究的其他阶段中明确引入多样测量以解决问题……但对多元方法策略更广泛的影响，包括多元方法的运用给分析资料、给写作和评审待发表的文章，以及给合乎道德地进行研究等所带来的各种新的挑战，还需要进行更详细的考察。（Brewer and Hunter，1989：21）

在本书第3章,我们还将依据Patton(1990)最早提出的分类,明确地将不同的哲学路径运用到单一研究项目的不同阶段(问题/假设的确定、资料收集和研究操作、分析和推论)。实际上,这一运用将成为本书其他部分的组织框架。

演变过程的不断发展

在过去的30年里,这种以应用混合方法和混合模型研究为方向的演变过程,以越来越快的速度发展。其原因在于:①大量新方法论工具的引进(既有定量方法的工具,也有定性方法的工具);②新技术的快速发展(如电脑的硬件和软件)使这些方法论工具的应用变得更为容易;③社会和行为科学领域跨界交流的与日俱增。

对不同方法论路径的分类研究

表1.1总结了三种主要的研究类型:单一方法研究、混合方法研究和混合模型研究。每一种研究的基本类型又都可以细分为不同的亚类型。我们将在本节对这些不同的方法论路径进行简单的界定,而关于它们发展和应用的更多细节将在随后的第3章中进行讨论。

单一方法研究

单一方法研究指的是"(方法论)纯粹主义者"所采取的在单一的主导范式内所展开的研究。显然,对其进一步的分类将包括纯粹的定性研究和纯粹的定量研究,而读者想必不会对此觉得奇怪。在本书第2章,我们将提供一些关于这种纯粹设计的案例。这种纯粹设计在社会和行为科学领域已经越来越少见了。

混合方法研究

混合方法研究指的是将定性研究路径和定量研究路径结合起来应用于研究方法论之中的单一的研究或多阶段的研究。在表1.1的展列中,这些方法可进一步区分为五种具体的设计类型。

表1.1所示的所有混合方法设计都使用了三角测量工具。这种三角测量工具从Campbell和Fiske(1959)的开创性研究中发展

而来。他们运用了不止一种定量方法来测量一个心理特性,并将这种技术称为“多元方法—多元特性矩阵”。Denzin(1978)则描述了四种不同的三角测量法类型,包括资料三角测量、研究者三角测量、理论三角测量和方法论三角测量。方法论三角测量是指,在对同一项研究或不同的补充研究之中同时使用定性和定量的方法和资料对同一现象进行研究。在一本关于评估方法的名著中,Patton(1990)列举了这四种三角测量的大量案例。

Creswell(1995)通过以下区分来界定表 1.1 所示的 4 种混合方法设计:

- 顺序研究(Creswell 称之为两阶段研究):研究者首先进行研究课题的定性研究,然后再进行课题的定量研究,或者以相反的顺序进行研究。定性与定量的两个阶段相互分开。
- 平行/共时研究:研究者同时展开定性和定量研究。
- 平衡设计:研究者大致平等地同时使用定性研究路径和定量研究路径来解释所研究的现象。
- 主—次设计:研究者“在某一主导范式内展开研究,但同时总体上也包含了一小部分另一类范式的设计”(Creswell, 1995:177)。

如表 1.1 所示,我们还界定了第五种混合方法设计:

- 多层次路径设计:研究者在资料收集的不同层次上使用不同的方法。例如,在学生层次上使用定量方法进行资料分析,在班级层次上则使用定性方法进行研究,在学校层次上再度使用定量方法,在地区层次上又使用定性方法。

巧合的是,Miller 和 Crabtree(1994)所提出的一套他们称之为多元方法临床研究的“工具”,与 Creswell 所界定的研究类型非常相似。他们所列出的混合方法设计类型包括:同时设计(concurrent design)(类似于平行/共时研究)、嵌套设计(nested design)(类似于顺序研究)和融合设计(combination design)(对上述设计选择作出某些融合)。他们的知识贡献推进了 Stange 和 Zyzanski(1989)的工作,后者曾率先呼吁在医学临床实践领域融合定量研究方法与定性研究方法。

混合模型研究

我们在表 1.1 中所列举的混合模型研究就是 Creswell(1995)所界定的"混合方法论设计"。Creswell 对此是这样界定的:"这种设计代表着最高程度的混合性范式……研究者在研究的所有阶段或许多阶段将定性研究范式和定量研究范式的因素混合起来。"(Creswell,1995:177-178)

我们关于混合模型研究的界定与此有所不同:混合模型研究是实用主义范式的产物,是在研究过程的不同阶段将定量路径和定性路径结合起来的学术努力。也就是说,在研究的每个不同的阶段内可以单一地应用一种方法,例如定量地(实验型)研究设计,随后进行定性的资料收集,并在转化资料后进行定量分析。在这种研究中,定性资料将依据 Miles 和 Huberman(1994)所描述的"定量化"技术而转化为数据。

在研究的各个阶段之中,也可以如下运用多种方法:

- 研究设计,要求有田野实验和高密度的人类学访谈,同时以整合的方式展开研究。
- 资料搜集,同时将以数值回答的封闭型选项和开放型选项置于同一项问卷调查之中。(如 Tashakkori, Aghajanian and Mehryar,1996)
- 资料分析,包括对问卷调查的某一部分按照李克特量表(Likert scale)所进行的因素分析,也包括使用常数对比方法(constant comparative method)对理论上与李克特量表相关的开放式问题的文字回答进行分析(如 Glaser and Strauss,1967;Lincoln and Guba,1985)。

本书的其他部分还将进一步描述这些混合模型研究的不同类型(见本书第 3 章的表 3.1),并说明如何设计这些混合模型研究。

2

实用主义与研究策略的选择

在本书第1章，我们对旨在将定量方法和定性方法结合起来的“混合模型研究”进行了概述；在后面的第3章，我们将对这些“混合模型研究”展开进一步的讨论。而在本章，我们将探讨与单一方法研究相关联的各种概念和方法论问题。

这一章的具体目的是：①讨论研究问题、方法、范式的相对重要性，并对实证主义、后实证主义、建构主义以及实用主义作出进一步的区分；②说明实用主义为何是行为和社会科学领域使混合方法和混合模型研究之应用具有确当性的适宜范式；③说明传统的（单一方法的）研究程序中的基本问题和原则，并提供更多的细节来说明表1.1所示的方法论路径类型学的演变；④进一步介绍这些方法论路径间的相异与相似之处。

如果您对范式间的差异已经足够熟悉，或对范式探讨不感兴趣的话，则无需阅读本章的前两节，请直接跳到本章的后两节，它们讨论的是单一方法研究中的设计问题以及传统资料收集技术的类型学。

研究问题主宰，而非范式或方法为王

对于研究者而言，一个重要的哲学性问题是关于范式、研究方法与研究问题的相对重要性。Guba 和 Lincoln（1994：105）在其最新的著作中重申，在他们看来，研究方法相对于研究范式只能居于次级地位：“定性研究方法和定量研究方法对任何研究范式皆可适用。

方法问题相对于范式问题而言是其次的,因为我们所说的范式是一个基本信念体系或者世界观,它不仅在方法选择上引导着研究,而且在本体论和认识论的基本途径上引导着研究。"

尽管有如此说法,但 Guba 和 Lincoln 仍然将具体的范式与具体的方法联系在一起。例如,他们将后实证主义与这样一些方法论——"修正的实验性/操纵性的方法,批判的多重方法论,可证伪的假设,也许还包括定性方法"——联系在一起。另一方面,他们又把建构主义与他们指称的"理解(hermeneutical)的或辩证的(dialectical)方法"联系在一起。① (Guba and Lincoln,1994:109)

Howe(1988:13)对范式(以及范式—方法论的关联)的重要性作了这样的实用主义回应:

> 但是,为什么范式应该决定研究者在研究过程中的工作方式?这难道比根据众多启发来决定研究者从何处入手更重要吗?……避免这种"方法的暴政(tyranny of method)"(Bernstein,1983)——以认识压倒实践,以概念压倒经验——正是实用主义哲学的标志。

与 Howe(1988)的这一评论相承接的则是 Trow(1957:35)的告诫:"让我们结束在'参与式观察'与访谈之间的争吵……全力以赴,用我们之所拥有且为解决问题之所需要的、最丰富的概念工具和方法论工具来对付我们所面临的问题"。

无独有偶,Brewer 和 Hunter(1989:17)也表示,多元方法使研究者得以"在攻治某个研究问题时拥有一个囤积各种方法的武器库,各种方法之间可以取长补短,相得益彰"。

简明扼要地说,我们认为,实用主义者更看重的不是他们所应用的方法,也不是支撑这些方法的世界观,而是他们所研究的问题。优秀的研究者大多喜欢不拘一格地使用可行的方法论工具来处理所研究的问题,其所采信的正是实用主义的教义:"有用即可(what works)"(如 Cherryholmes,1992;Howe,1988;Rorty,1982)。对于大多

---

① 这种方法论(Guba and Lincoln,1994:111)包括:a. 通过研究者与研究对象之间的互动,引出并完善个别的理论建构;b. 通过参与者之间的理解技巧和对话,对理论建构加以阐释;c. 决定共识的建构。(根据 Kneller 的著作(1984:66),理解性哲学指的是"对理解的哲学研究",它被看作是"阐释性的"。理解性指的是对文章或文献的分析)

数专心致志于某一问题的研究者来说,相对于研究问题本身,方法只是次要的,而隐含的世界观除了在最抽象意义上加以议论之外,也很难进入其视野。虽然本书至今涉及的只是范式之争及其解决,但仍需重申,这种有关范式的考量在最终的分析中远不如研究者所试图回答的研究问题那样重要。(关于研究问题类型的详细阐述,见 Hedrick,Bickman and Rog,1993:23-24)

最优异的学者总是对其所提出的问题切思在心,而对所用之具体方法论或是方法背后之范式则洒脱不羁。例如,在方法论先驱人物 Donald T. Campbell 弃世不久,Gene V. Glass 在互联网上发文诔赞之:"方法必从之于问题。数十年之前,Campbell 即倡导三角测量之法——体认到诸法皆有不足,总需多法并用。"

范式之比较

表 2.1 展列了我们认为属于社会和行为科学领域所应用的四种主要范式——实证主义、后实证主义、实用主义和建构主义之间的差异之处。有关实证主义、后实证主义、建构主义的这些比较提炼自 Denzin 和 Lincoln(1994)、Lincoln 和 Guba(1985)、Guba 和 Lincoln(1990,1994)、House(1994)以及 Miles 和 Huberman(1994)等人的著述。关于实用主义的信息则摘自 Cherryholmes(1992)、Greene(1994)、House(1994)、Howe(1988)、Murphy(1990)和 Rorty(1982)的论说。(批判理论及其变体没有列入这一对比中,因为批判理论强调历史方法,而本文的关注点在于对现行现象的考察。)

有趣的是,理论家们在对比描述建构主义(或解释主义)与实证主义(包括后实证主义)时,通常不会将实用主义作为第三方纳入比较视野(如 Guba and Lincoln,1994;Smith and Heshusius,1986),尽管前两者的比较并没有穷尽范式的可能性。正如 Howe(1988:13)所指出的那样:"这看来是一个严重的纰漏,因为实用主义者对扳倒实证主义贡献很大,他们会断然拒绝在解释主义与实证主义范式之间作出非此即彼的被迫选择。"

表 2.1 社会和行为科学领域所应用的四种主要范式之比较

主要范式	实证主义	后实证主义	实用主义	建构主义
研究方法	定量	定量为主	定量+定性	定性
推论逻辑	演绎	演绎为主	演绎+归纳	归纳
认识论	客观论 认识主体与认识对象是二元关系	修正的二元论 研究发现可能是客观的"真实"	客观论和主观论并存	主观论 认识主体与认识对象是不可分割的
价值观	研究是价值中立的	研究含有价值倾向,但是可控的	在阐释结果时,价值有很大影响	研究受到价值的限定
本体论	天真的现实主义	批判的或超验的现实主义	承认外在的现实,选择最能产生所欲之结果的解释	相对主义
因果联系	结果之前或同时,必有真实的原因	社会现象之间存在某种规律的、合理的、稳定的关系,但无法完全知晓。原因在因时变化的概率意义上是可辨识的	可能存在因果联系,但我们永远无法将之确定下来	一切事物都会同时相互依存,不可能区分原因与结果

拒绝"非此即彼"

在表 2.1 中,突显了实用主义拒绝在方法、逻辑及认识论方面对实证主义(包括后实证主义)与建构主义作出非此即彼的被迫选择。无论对哪一方,实用主义者都拒绝不相容论的"非此即彼"命题,而对两种观点都张开双臂表示欢迎。Nielsen(1991:164)提出,实用主义是一种"相反的、批判性的哲学",它反对占主流的体系哲学,讥讽"非此即彼"之类的形而上论断。

关于研究方法,通贯本书我们都在探讨实用主义者同时使用

定性方法和定量方法的倾向。尽管表2.1指明后实证主义者可能也会采用定性方法，但不要忘记第1章中已经讨论过的“方法论正确性（methodological correctness）”。在选择定性方法论或定量方法论时，后实证主义者由于对因果联系及内部效度关注，往往会选择实验型设计（或其变体，如准实验型设计、追溯设计或过程比较设计）（如 Cook and Campbell，1979）。

与之相似，建构主义者则青睐于他们自己的方法，恪守方法论取向上的门派之别。例如，Denzin 和 Lincoln（1994：6）在《定性研究手册》一书中，提出了以下他们认为泾渭分明的不同类型：

> 上述五点不同，反映了对不同的研究风格、不同的认识论以及不同的表达方式的信念。每一种研究传统都有其作为流派的一套门规……定性研究者采用民族志散文、历史陈述、第一人称描述，另一些人还采用照相、生活史、小说化的事实、传记和自传材料等。定量研究者则采用数学模型、统计表和图表等。

然而，实用主义者却认为两类方法都有其用处，Denzin 和 Lincoln 所罗列的一系列令人眼花缭乱的定性方法和定量方法皆为我所备，尽为我所用。至于究竟是用定性方法还是定量方法（或两者都用），则取决于当下所面对的研究问题以及研究之轮所在的阶段。

研究之轮：归纳逻辑和演绎逻辑的统一

对任何问题的研究无论何时都会处在推理过程循环圈的某一点上，这个循环圈常常被称作研究之轮、推理之链（如 Krathwohl，1993），或科学方法论之轮。这个转轮可以看作是从基础结果（即事实、观察所得）经过归纳逻辑，变成普遍性推断（或理论），然后又从这些普遍性推断（或理论）经过演绎逻辑，变成待检验的假说或对特定事件和结果的预测。图2.1形象地表达了这种推理链。

任何实存领域的研究在其结束之前至少要在这个圈里完成一次循环。研究可以从这个循环圈的任何一点开始：一些研究者从理论或抽象概括开始，一些研究者则从观察开始。在许多研究

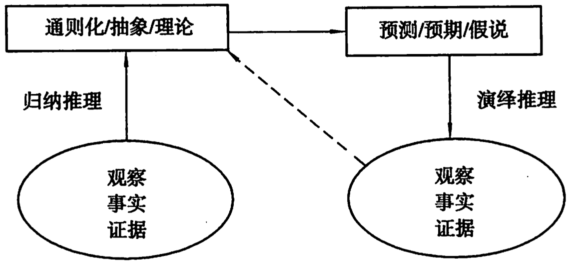


图 2.1 研究之轮(科学方法论之轮)

报告中,研究者最初(即在一篇期刊论文的导论部分)都试图通过归纳的方式在以往研究的基础上建立一个概念(理论)的框架。这样获得的概念(理论)框架即作为设计研究过程的基础。无论研究者从哪里(事实或理论)开始其研究,由于总有一个需要得到满意回答的问题,一个研究项目便往往由此展开,并且在某种程度上至少经历一次这个循环。

在研究过程的某些时候,两种类型的推理和方法可能会同时得到使用。当此之时,我们就有了表 1.1 所列举的、在研究阶段中多元并用的混合模型研究。实用主义者认为,在研究一个待解问题的过程中,他们会有选择地使用归纳逻辑或演绎逻辑。

认识论相对主义:主观观点和客观观点的同时使用

不相容理论黑白分明的对比(不是二元论,就是一元论;不是客观论,就是主观论)再度受到实用主义者的质疑。在实用主义者看来,科学研究不是"形式主义的"(formalistic):对于问题研究过程的认识论取向,研究者既可以是主观论者又可以是客观论者。就此而言,理当将此二者视为一个连续谱,这远比将之视为对立的两极更为合理:在某些时候,研究者可能更为"主观"一些,但在另一些时候,研究者可能就更为"客观"一些;在某些问题上,知者与被知者必须是互动的,而在另一些问题上,研究者也许更容易与被研究者保持距离。

如果允许研究者同时使用定性研究和定量研究的方法论工具,那么主观论与客观论的兼容并蓄就是不可避免的。

价值的作用(价值论)

一如表 2.1 所示,实证主义者认为研究是价值中立的;建构主

义者则认为研究是受价值限定的；后实证主义者认识到价值在研究中有很大的影响，但他们同时也相信，可以将价值对研究结果和解释的影响程度加以控制。

第1章已经提到，后实证主义者承认事实中渗透着价值，也渗透着理论（Reichardt and Rallis，1994）。尽管承认了这一说法（在很大程度上也正是因为这一承认），后实证主义者（如 Cook and Campbell，1979）仍付出了大量的努力来发展使其结论的内外效度都得到提高的方法。这些旨在提高结果及其解释之效度的方法，表现了后实证主义者试图降低他们的个人价值取向及其忠诚于特定理论立场的影响。

在实用主义者看来，在展开研究操作以及得出研究结论之际，价值扮演着很大的角色。然而，他们认为没有必要为此心理负担过重。正如 Cherryholmes（1992：13-14）所言：

> 对于实证主义者而言，价值观念以及对人类行动、互动的想象，优先于对描绘、理论、解释和叙述的寻求。实用主义研究受着预期结果的驱动；关于研究什么以及如何进行研究，实用主义的选择受制于我们在最宽泛的意义上所设想的研究方向……根据其已有的知识以及所欲获得的结果，实用主义者会在左右权衡中确定研究的方式和对象，以及所要做的工作。

就此而论，实用主义者是在其个人价值观体系的指导下决定他们想要研究的对象，也就是说，他们研究其自认重要的研究。随之，他们用适合其价值观体系的方式研究特定的主题，包括使用他们认为最适于解答该研究问题的各种变量和分析单位，他们也希望他们的研究结果与其价值观体系相一致。对研究者用以展开其研究之方式的这种解释，似乎就是描述社会和行为科学领域的研究者在从事其研究、尤其是有着重要社会影响的研究时所实际采用的方式。

关于本体论和因果联系的思考

与实证主义/后实证主义、建构主义以及实用主义观点相关联且最具争议的问题，牵涉到本体论和因果联系（见表2.1最后

两行)。Guba 和 Lincoln(1994:109)认为,有关现实之本质(本体论)的观点分歧构成了实证主义、后实证主义、建构主义之间的定义性区别,"建构主义这个词是指一种替代性的范式,其叛逆性的设想就是舍弃本体论现实主义,而转向本体论相对主义"(原文为斜体)。Guba 和 Lincoln(1994)、Miles 和 Huberman(1994)界定了下述几种类型的现实主义:

1. **天真的现实主义。**存在着一个客观的、外在的现实,对此,科学研究也可给予确认(Guba and Lincoln,1994:111)。如表 2.1 所示,这一立场与实证主义者联系在了一起。

2. **批判的现实主义。**存在着一个客观的现实,但只能不完全地、以概率论的方式加以理解(Guba and Lincoln,1994:111)。这种立场是后实证主义者所持有的。

3. **超验的现实主义。**这一派别相信社会现象存在于客观世界之中,这些现象之间存在一些"规律性的、合理的、稳定的关系"(Miles and Huberman,1994:429)。这一表述是后实证主义立场的另一种说法。

4. **本体论现实主义。**存在着多种多样的社会现实,它们是人类智慧的产物,会随着建构者的不同而不同(Guba and Lincoln,1994:111)。这是建构主义者的立场。

关于现实的本质,实证主义/后实证主义和建构主义之间最大的分歧在于是否存在着客观的外部世界:实证主义/后实证主义认为存在着这样一种客观的现实(关于这方面争论的探讨,请见 Greenberg and Folger,1988:第一章),而建构主义则认为存在着的只是多元的、主观的现实。

而关于现实,实用主义的观点由两部分组成:

1. 存在着一个独立于我们思想的外部世界(Cherryholmes,1992:14)。因此,在外部现实的存在问题上,实用主义者赞同实证主义者/后实证主义者。

2. 另一方面,实用主义者认为,"真理"不可能一劳永逸地得到确定,他们也不能肯定对现实的某一种解释是否就比另一种解释更好。在 Cherryholmes 看来,实用主义者选择某一种解释而非

另一种解释,"只是意味着这一种方法比另一种方法能更好地产生预期的或想要的结果"(Cherryholmes,1992:15)。

就实用主义者关于真理的观点而论,Howe(1988:14-15)进一步解释道:

> 对于实用主义者而言,"真理"是一个规范性概念,就像"善"一样。"真理即能用者",这最好不要看成是一个理论或定义,而应视为实用主义者尝试对真理属性所作的生动的解释,特别是表明知识的获得并不能完全从偶然性的观念、利益和项目中抽象出来。

关于因果关系的看法则依循着这些本体论方面的区分而衍生出来。简言之:

1. 后实证主义者相信,社会现象之间存在着一些规律性的、合理的稳定关系,可以在不完整或概率论的意义上加以理解。尽管通过某些事件(预测变量、"原因"等)而对另一些事件(标准变量、"效果")所作出的预测不可能达到100%的准确(1.00的概率),然而预测的准确性却可以随着时间以及有效预测的确认而得到改进。
2. 实用主义者认为,可能确实存在着因果关系,但是我们永远无法将之完全确定下来。
3. 建构主义者认为,所有的实体都是同时相互塑造对方,因而不可能将原因和结果区分开来。

因此,对于理解因果联系的可能性,从后实证主义者到实用主义者再到建构主义者,悲观主义色彩越来越浓厚。例如,House(1994:18)在回应Cherryholmes的实用主义观点时,阐述了科学的现实主义(后实证主义)观点:

> 对解释作出判断的标准在学科与学科之间、在学科的内部都有所不同,且不断变化,但是,它们总与物质现实有着某种联系。学者们在提出新的观点时——也许使用新的故事、隐喻,他们对自然世界创造出了新的解释。然而,他们并没有创造世界,世界早已存在。这也不意味着他们的解释是最终"正确的",不会出现任何终止的红灯,而且也不必要。存在着的只是

为学者共同体所讨论的更好的解释,这些学者接触着物质现实,并常常以语言、概念和形象去解释这一物质现实。就此而论,实用主义者对解释现实世界的可能性似乎过于悲观。

在现实的本质与我们理解因果关系的能力上,我们对后实证主义者与实用主义者之间的争论感到有些相互矛盾。两种范式都认为存在着一个外在的现实世界,我们应当探求其中的因果关系。

同时,他们又都赞同事实不能充分证明理论的原则,换言之,任何一套资料集都可以用多种理论加以解释。(Reichardt and Rallis,1994)

一方认为,我们应当永不止息地去寻求对现实和因果联系的更好解释;另一方却认为,我们应当赞同那些与我们自身的价值观更相近的对因果联系和现实的解释,因为我们永远不能彻底地理解因果联系和现实。正是由于任何研究结果都会有多种解释,所以,这就归结为或者选择"更好的"解释(后实证主义者),或者选择更接近研究者之价值观的解释(实用主义);而由于研究者设计了研究课题并建构了其操作性定义,那么,选择"更好"的解释与选择更接近研究者价值观的解释,往往是同一回事。

这两种观点最基本的区分还涉及对于发现真理是否持有乐观主义态度的问题。就此而论,我们的立场最好称为谨慎的乐观实用主义。

实用主义的天生魅力

考虑到所有的因素,实用主义看来是证明混合方法和混合模型的应用具有确当性的最佳范式(Howe,1988)。实用主义的魅力在于:①它为我们提供了可以从哲学上将混合方法和混合模型设计的应用包容在一起的哲学范式;②它摆脱了应用那些曾引起了无尽(也多为无用)的讨论和争辩的形而上概念(如"真理"、"现实");③它提出了一种非常可行和实用的研究哲学:研究你所感兴趣的和对你有意义的问题,应用你认为合适的不同方式去研究这些问题,并以能为你的价值观体系带来积极后果的方式使用这些研究的结果。

与单一方法有关的设计问题

在单一方法时期(表1.1所说的第一阶段),研究中所用的方法与研究者所行的科学范式之间存在明确的联系。如第1章所指出的,方法论纯粹主义者坚持不相容理论,认为定量研究者不应该使用定性方法,反过来,定性研究者也不应该使用定量方法。

传统的单一方法研究不仅主张在两种主要路径中只能遵奉其中之一,而且主张在资料收集的类型以及资料分析的程序,乃至于依据资料分析而进行推理的类型上,也只能恪守迥然不同的两条路径中的一种。但下面的讨论表明,这两种单一方法实际上存在着许多相似之处。

单一方法之原型的比较

单一方法设计的原型有两种:①实验室实验,其特点是有一个变量之发生受到操纵的受控研究环境,并且是对预先建构的假说进行确证性检验(confirmatory investigations);②描述型案例研究,其特点是有一个变量没有受到操纵的自然环境,同时,所进行的是探索性研究(exploratory investigations)。

这些传统的研究技术往往会在三个主要维度上相互区别:第一个维度涉及是否存在着对自变量的操纵;第二个维度关系到研究的环境(自然环境对受控环境);第三个维度涉及是否在研究发现的方向上有事先的假说和(或)预测。

关于单一方法路径的争论,大多围绕前两个维度展开,更具体地说,是集中于操作和控制的问题展开。Kerlinger的最大最小控制原则对这些问题作了很好的总结。框2.1即是对这一原则的介绍。

两种单一方法的原型真是对立的两极吗

正如我们前面所指出的,在讨论研究设计的问题时,与其采信非此即彼的二分法,毋宁将这些问题看成一个连续谱要来得更为有效。例如,将纯定量研究视为连续谱的一端,而将纯定性研究视为另一端,那么,这两个端点的中间就会存在着极其丰富的不

框 2.1　最大最小控制原则(the MAXMINCON principle)

在讨论研究设计时,Kerlinger(1986)将不同的研究设计问题总结为他所命名的最大最小控制原则(the MAXMINCON principle)。这个原则的三个组成部分(最大化实验差异、最小化错误幅度以及控制外部变化)可以理解为进行研究所必需的步骤。

最大化实验差异(experimental variance)是指有必要允许不同的组群(或自变量的不同程度)有足够的差异,从而确保产生效果。例如,如果两种教学方法之间没什么不同,那么,它们的效果差异也就是不可测度的,即便它们也许确实非常有效。用 Kerlinger 自己的话讲,"设计、计划与开展研究,尽可能使实验的环境不同"(Kerlinger,1986:287)。

最小化错误幅度(error variance)为检测组群或变量关系之间的差异提供了支持。那些随机的波动如同在你试图听清谈话(信号)时却一直回响在耳边的随机噪音:噪音的波动越大,那么就越难听清信号。错误幅度就是反应、行为和(或)测量中的随机波动所产生的结果。测量的不可靠性是这种错误幅度的一个来源,个体差异则是另外一个来源。通常,随着观察数量(样本规模)的增加,个体差异的来源也会随之增加,而随机错误也就相互抵消了。

控制外部变量意味着所有会影响因变量的其他竞争变量都应该被排除(加以控制)。没有这种控制,研究者就不能对研究中的主要自变量展开推论。例如,如果教学方法之外,性别问题也与学习成绩有关联,那么,学生组群之间的差异也就不能归结于教学方法。

尽管最大最小控制原则可能看上去更适用于实验方法,而不是非实验方法和定性方法,但是,它仍然适用于所有的研究。研究者对某种关系的研究发现所持有的信任程度(即定量研究者所谓的"内部效度"),取决于研究者排除其他竞争解释的能力。

同的研究设计类型。对于连续谱的偏好体现了将研究设计问题看成是灰色梯度而非黑白分明的重要意义。事实上,假如你能作出这样一个主要的假定:定性研究和定量研究的目的都包括回答如何和为何的问题(关于因果关系的讨论,见 Shadish,1995),那么,这两种研究设计的原型(实验研究与案例研究)是否真的彼此不同,并不是清晰的。

如表 2.1 所示，关于建立因果联系的可能性问题，从实证主义者到建构主义者之间各派哲学立场纷然杂陈，不一而足。然而，就我们看来，在实际从事研究时，许多在定性案例研究传统中工作的研究者同样也在试图回答因果关系的问题。如同许多定量取向的方法论学者也采用了一些建构主义的原则一样，许多定性取向的方法论学者也相信因果联系是可能的（有关综述见 Shadish，1995），尽管由于在自然环境中难以实现控制，他们对建立因果联系要更为谨慎一些。

我们不相信大多数研究者（无论是使用定性方法、定量方法或者是混合方法）会满足于仅仅指出"发生了什么事情"，我们相信他们都想解释事情是怎样发生以及是为什么发生的。例如，在框 2.2 中所总结的 Whyte（1943）的经典案例研究中，他描述了在社会结构、组织结构和职业晋升等不同变量之间的因果联系。Yin（1993）令人信服地指出，案例研究不应仅停留于描述性问题（谁、什么、哪里、多少），还应对如何以及为什么的问题作出回答（例如，按 Yin 的看法，可以进行实验研究和历史分析）。

框 2.2　定性案例研究的一个案例

William F. Whyte 所著的《街角社会》（*Street Corner Society*），几十年来始终是社区社会学研究的推荐读物。这本书是描述性案例研究的经典案例。书中追溯了历时性人际事件的先后发生过程，描述了此前的研究很少涉及的亚文化主题，并且发现了一些关键性的现象——如低收入年轻人的职业晋升以及他们能否突破邻里纽带。这项研究受到了高度的重视，尽管它是一项单一案例研究，只描述了一个街道（cornerville），且距今已有 50 年之久。具有吊诡意味的是，这本书的价值在于它对个人行为、组织结构以及邻里社会结构的问题所作出的通则化。后来的研究者的著作中总能找到一些"街角社会"的影子，尽管他们研究的是不同时段的不同街道社区。

来源：摘自 Yin（1993）。

如果说定性研究者对研究因果联系也抱有研究兴趣，那么，对比研究纯粹定量研究与纯粹定性研究以考察它们之间究竟有何不同或相似，显然是有益的。在此，我们将在这样三个维度上对实验室型研究与描述性案例研究进行简要的比较：对变量的操纵，研究

所在的环境类型(自然环境对受控环境),以及研究的性质(确证性研究对探索性研究,或预设性研究对非预设性研究)。为建立讨论的平台,提供可与 Whyte 的研究(框 2.2)相比较的研究案例,框 2.3 提出了一个传统的、确证性的定量研究(Latane and Rodin,1969)。

框 2.3 传统的、确证型的定量研究的一个案例

Latane 和他的同事在 1960 年代后期开展了一系列实验来进一步探讨所谓"旁观者介入"的现象,这种现象所涉及的是陌生人帮助不幸者的意愿问题。由于一个受到广泛关注的案例——一个年轻妇女遭到谋杀,而她的 38 个邻居却袖手旁观、无动于衷,实验社会心理学家由此开始对这一问题给予关注。

研究者提出了一种理论,预测说"旁观者"的数量越多,受害者获得帮助的可能性也就越小。他们的推论是,不介入的行为是一种从众性行为:个人是根据其观察他人所得到的提示来对受害者的危难作出反应的;如果其他人都不提供帮助,这就会使个人将不介入理解为妥善的行为。

在一项实验中,Latane 和 Rodin(1969)设计了一个以"遇险的妇女"为中心的实验情境:一名女性实验人在一个打开的卷帘后假装发生了一个"事故",使被研究者不能看见但能听见。在研究中,被研究者听到的是一盒关于这名女性爬椅子、摔倒、然后求救的录音带。

一种实验情形是,被研究者是该情境中唯一的他人。另一种实验情形是,情境中有一些与实验人串通好故意对"受害人"无动于衷的"配角"。理论上作出的预测由此得到印证:在"单独"的情境中,70% 的被研究人对"遇险的女性"作出了反应,而在有"配角"的情形下,只有 20% 的被研究人伸出了援手。

单一方法之原型的比较:对变量的操纵

实验研究与案例(或田野)研究的传统区分,强调实验研究有对变量的操纵,而案例研究则没有对变量的控制。事实上,从事案例研究的研究者往往会试图使其观察案例中的差异实现最大化(最大最小控制原则中的最大化原则)。

实验研究往往还可以区分为实验室环境、田野环境和自然环境三类实验研究。在前两种实验类型中,实验者本身是自变

量变化的主要动因；而在第三种实验类型中，自变量的变化是其他力量（社会制度、法律、自然等）作用的结果。研究者的角色就是对发生了明显变化的群组和（或）环境与未发生明显变化的群组和（或）环境进行比较研究。案例研究中的"处理"在经由这类"自然的"控制而发生时，研究者也通常会使用定性的抽样方案（例如极端抽样、最大变化抽样），以便使其研究的群组之间的差异能够凸显出来。例如，研究者会对高效率的学校与低效率的学校进行对比研究，以确定这些学校对学生学习所造成的不同影响（如 Brookover, Beady, Flood, et al., 1979; Teddlie and Stringfield, 1993）。

对差异极大的群组进行案例研究，不同于使用已经经过不同的实验处理的群组进行实验室研究或田野实验研究，关键在于一个需要控制自然，而另一个则需要研究者进行控制。这样一种方法论上的差别看来并未构成一个严格的二分法以截然区分案例研究与实验研究，而是构成了一个连续谱，横跨了从研究者在"处理"案例时作为变化的动因到研究者对变化没有任何控制的不同情形（第 5 章的一个延伸案例进一步说明了不同的研究者在研究同一现象时在"处理"方法上的不同）。

单一方法之原型的比较：自然环境对受控环境

实验研究与案例研究的一个通常（或传统）的区别，是认为实验研究发生在受控环境之中，而案例研究则发生在自然环境之中。从研究设计的角度来看，这种区分的本质在于对外部变量的控制（最大最小控制原则中的控制原则）以及对处理过程中误差幅度的降低（最大最小控制原则中的最小化原则）。

较之于自然环境下的研究，受控环境下的研究更有可能降低外部变量的影响以及误差的幅度。然而，总是还存在着一些无法控制的变化——即使是最精心设计的实验也永远不可能完全消除这些变化。事实上，在受控环境下所进行的研究会导致一些不同类型的外部变量，例如"实验者偏好效应（the experimenter bias effect）"（Rosenthal, 1976）。在这种效应中，实验者的预期会以某种方式传递给他们的研究对象，进而影响到研究对象的行为。

另一方面，设计良好的案例研究也可以包括对重要变量的

"控制",而相对避开外部变量和误差幅度的影响。例如,在对高效率的学校与低效率的学校进行比较的案例研究中,研究者们在研究设计中已经对下列变量进行了控制:入学学生的社会经济地位、学校的族群组成、对象学校所在的地域、对象学校所在的社区类型、作为研究对象的年级、讲授和考试的课程类型,以及其他一些因素,等等(如 Chrispeels,1992;Hallinger and Murphy,1986;Stringfield and Teddlie,1990;Teddlie and Stringfield,1993)。

显而易见,这一讨论涉及研究的内部效度和外部效度这个一般性的问题,我们在第4章和第5章中会对此展开更详细的讨论。在此,我们想指出的是,在控制外部变量和误差幅度这样的重要问题上,自然环境与受控环境的区分也不是截然的两分,而是一个连续谱。进一步说,实验可以在自然环境下开展(田野实验),而案例研究也可以在受控环境下开展(临床案例研究)。

单一方法之原型的比较:确证型研究对探索型研究

传统上对实验研究与案例研究的区分还认为实验研究在本质上是确证型的,包含着事先假说的使用;而案例研究在本质上则是发现型的,不涉及对假说的检验。但是,正是在这方面存在着大量的反例:案例研究中有假说,而实验研究中却没有假说。

这一问题与表2.1中所显示的研究之轮有关。在这个研究之轮中,归纳推理和演绎推理被视为在现象研究的整个过程中相互补充、相互替代的组成部分。在这里,实验研究和案例研究之间的区分再一次被模糊了,每种方法都能够适合于确证型研究或探索型研究。

因此,如果研究者关心的是因果联系,那么,在我们考察过的三个维度上,定量方法(实验室型实验)和定性方法(案例研究)这两种典型就没有什么不同。Yin(1994)最近在讨论定性研究和定量研究的共性时,也得出了同样的结论。他在一些定性研究和定量研究的示范案例中,开列了两者间的四个共同点:对所有证据的详尽描述和研究,对对立假说的持续关注和检验,寻求有重要意义的结果,以及针对研究对象的研究专长。

社会和行为科学领域中传统资料收集方法的类型学

表 2.2 中对行为和社会科学领域的传统研究设计进行了归类整理。我们将此表称为“资料收集方法”而不是研究设计的分类，其缘由我们将在本节的后面作出解释。在此表中，研究设计的常见类型是依据上面所提出的三个分析维度所概括的。表中的信息也告诉我们，传统设计主要可以划分为三大类别：

1. 第一类设计是控制环境以及（或）控制变量，旨在检验尝试性的预测或假设（确证型研究）。这一类别的设计包括实验型实验、单一对象研究和田野实验。

2. 第二类设计是既不控制环境也不控制变量，也许是确证型研究，也许是探索型研究。这一类别的设计包括问卷调查、相关性/档案性研究、因果比较研究、历史研究和案例研究。

3. 第三类设计是既不控制环境也不控制变量，且聚焦于探索性问题。这一类别的设计包括描述性研究和民族志研究。

尽管这一分类法突显了一些有趣的对比，但它并不是一种令人满意的、能够有效地将社会和行为科学领域所有不同的设计类型区分开来的方式。这种分类法的第一个问题就在于设计类型的数量上，表 2.2 中有 12 个。Brewer 和 Hunter（1989）构建了一个相似的分类法，但研究类型的数量是 4 个。在 Yin（1993）的分析中，设计类型则是 5 个。不同的研究领域有不同的特殊属性，不同领域的研究专著因此在其分类法中包含了不同的因素。

第二个问题是，这种分类法的基础是我们在本章前面部分的讨论中所提到的三组对应的二分法，而它们却有点“模糊不清”。例如，在 12 种研究设计中，有 4 种对确证型维度和发现型维度进行了考察。尽管表 2.2 对环境类型以及是否存在控制作出了更清晰的区分，但上述讨论表明，实验研究和描述性案例研究要比这一表格所显示的更具相似性。举例而言，确证型案例研究即具有试图控制自然环境下某些外部因素的设计特点。

表 2.2 中分类法的最后一个问题，涉及研究设计在跨越学科界线时所携带的方法论“辎重”。这种“辎重”的出现是由于不同

学科在研究设计上都有一些几十年不变的先入为主的观念。其中,一个很好的案例就是案例研究。案例研究可以是确证型的吗? 以实验心理学背景进行写作的 Yin(1993)认为是的,但许多人类学和教育学民族志研究者却会称否。

在本书第 3 章,我们为混合模型设计提供了一个更具逻辑性的分类法,单纯立基于这些设计的内在特性。我们力求在这种混合模型研究的分类法中避开那些渗透着价值观的术语(见表3.1)。

我们此前将表 2.2 中的分类法称为资料收集的方法,而不是研究设计或研究类型的系列方法,但其他一些标准教科书中所采用的却是后一种名称(如 Gall, Borg and Gall, 1996)。我们之所以采用不同的命名,是出于两个原因:

1. 表 2.2 沿袭的是传统的分类法,不同的设计类型之间内在地存在着重叠,学科间的界线也模糊不清,而我们更倾向于将研究设计以无重叠的方式区分开来的分类法(将在第 3 章进行讨论)。

2. 事实上,表 2.2 中的类型所涉及的基本上是不同的资料收集路径。这些路径来源于所有的社会和行为科学:实验室型研究来自于心理学,单一研究对象设计来自于学校和教育心理学以及特殊教育,田野实验来自教育学和评估学,问卷调查来自社会学,相关性和档案性研究来自社会学等多个领域,(事前的)因果比较研究来自评估学、教育学和社会学,历史编纂来自于历史学,案例/田野研究以及民族志研究主要来自人类学(社会学也对此有贡献),而描述性研究实质上几乎是所有社会和行为科学领域中都会有的。

这些路径代表了这些不同学科中传统的、规范的资料收集方法。正因为这样,它们与这些学科中的研究设计发生了混淆:心理学方法与实验研究、社会学方法与问卷调查、人类学方法与民族志研究、历史学方法与历史编纂以及因果比较与评估研究等。

事实上,社会学家(或任何其他领域的专家)没有任何理由不可以在从事的社会学研究(或其他类型的研究)中,使用实验、问卷调查、民族志研究、历史编纂、准实验研究以及所有其他类型的资料收集方法。许多顶尖的研究者经常使用本不属于他们学科的研究设计来解决研究问题。在本书第 5 章,我们还将进一步对这些特定的资料收集方法进行讨论。

表 2.2　在社会和行为科学领域关于传统资料收集方法的类型学

资料收集方法	环境		变量控制		研究取向	
	受控的	自然的	有	无	确证型	发现型
实验室实验	×		×		×	
单一对象研究	×		×		×	
田野实验		×	×		×	
问卷调查		×		×	×	×
关系研究		×		×		
预测研究		×		×	×	×
档案研究		×		×		×
因果比较		×		×	×	×
历史研究		×		×	×	×
案例/田野研究		×		×	×	×
描述性研究		×		×		×
发展性研究*	×	×		×		×

* 发展性研究(developmental research)包括纵向研究、横向研究、短时段纵向研究及相似的研究(见 Salkind,1997)。尽管这些方法被划分为描述性研究或问卷调查,但我们认为它们应当划分为一个独立的类别,因为研究者在这些研究中可能混合采用从定性案例研究到相关性/预测性研究以及准实验研究的多种方法。

3

混合方法和混合模型研究的研究设计问题

本章论说了关于混合方法、混合模型研究的诸多方法论问题，提出了一种有关混合模型研究的分类法，而这一分类法也成为了本书后面的内容的论述框架。

单一方法的不足：多样化的缺陷

在实用主义者令人信服地提出了相容理论之后，范式与方法之间的联系不再是不可动摇的了。逐渐地，许多社会和行为科学研究者也开始接受其他学科的方法论取向。在这一过程中，对不相容理论的批判使单一方法研究设计的各种不足彰显了出来。例如，Brewer 和 Hunter(1989:16-17)在如下的引言中将单一方法设计描述为“多样化的缺陷”：

> 各种社会科学方法不应被视为相互排斥、非此即彼的选项……我们的单一方法设计可能会有缺陷，但幸运的是，不同的单一方法设计的缺陷并不一样。缺陷的多样化使得我们能够将这些方法结合起来加以使用……进而使其失误与缺陷得到相互补偿。

与此相似，Cook 和 Campbell(1979)在就研究结果之效度的威胁因素进行讨论时，指出单一方法在测量基础性思维建构物(underlying constructs)时存在着缺陷。这些定量方法论的名家学者将“单一方法偏执”看成是对假定因果(也就是据认为的因果)之建构效度的一种威胁。他们论证说，如果仅仅用一种方法来测

量一个思维建构物,那么,就会很难将此一建构物与其特定的单一方法的操作性定义区分开来。以只使用纸笔记录的回答来测量态度为例,他们质疑道:“如果单独测量的‘个人的私下态度’与‘非叙述性的纸与笔回答’相反,那么研究者是否还能对之进行检验呢?”(Cook and Campbell,1979:66)。这种对单一方法的批评继承了本书第 1 章中所提及的 Campbell 和 Fiske(1959)的早期著作,后者确信在估测心理特征时需要使用多重定量方法。

与混合方法研究相关的设计问题

在混合方法时期(表 1.1 所指称的时期Ⅱ),逐渐出现了大量同时包含着定性取向与定量取向的研究设计。这些设计在许多领域得到了发展,这在很大程度上是三角测量法日益流行的结果。

三角测量法概念的重要性

三角测量法的概念是一个智力性的楔子,它最终钻破了纯粹单一方法的支持者们在方法论上的霸权格局。这一过程经历了一系列演进,包括如下阶段:

1959 年,Campbell 和 Fiske 在其实验心理学的著作中提出了“多元特性—多元方法矩阵”(multitrait-multimethod matrix),主张应用一种以上的定量研究方法来测量某一种心理特性。他们提出这一矩阵是为了确保研究中出现的差异确实源于所研究的特性本身,而非源于所应用的方法(Brewer and Hunter,1989;Creswell,1995)。

我们已经在表 1.1 里区分了使用单一资料来源的纯粹定量设计和使用多元资料来源的定量设计。正是 Campbell 和 Fiske 的著作使得同一研究课题中应用多元定量技术的做法流行起来。

1978 年,Denzin 在一本关于社会学方法的著作中使用了“三角测量法”(triangulation)一词。三角测量法最初是指一种使用两点(及其角度)来探寻到第三点之未知距离的测量/航海方法。Denzin 所谓的三角测量法是将多种资料来源结合起来,对同一种社会现象进行研究。他提出了 4 种基本类型的三角测量法:资料三角测量法(data triangulation,在研究中使用多种资料来源),研究者三角测量法(investigator triangulation,使用不同的研究者进行同一研究),理论三角测量法(theory triangulation,使用多元视角来解释研究结

果),以及方法论三角测量法(methodological triangulation,使用多元方法来研究某一研究问题)。

1979年,在行政学领域从事著述的Jick,又从一种方法的弱点可以用其他方法来弥补的角度讨论了三角测量法问题。Jick还论述了同类方法内的三角测量法(within methods triangulation,如多元的定性路径或多元的定量路径)及不同方法间的三角测量法(across methods triangulation,如同时使用定性路径和定量路径)。

Patton(1990)在一本著名的关于定性评估和研究方法的著作中,描述并举例说明了三种三角测量法:融合定性和定量资料(不同方法间)的三角测量法,比较研究多元定性资料来源(同种方法内)的三角测量法,以及来自多元观察者之多元视角(定性资料的不同分析者之间)的三角测量法。

在对教育或社会服务项目进行评估时,Patton详尽地阐述了使用多元定性资料来源的三角测量法。这通常需要同时使用访谈法及观察法。Patton和Jick的论著引导了同一研究领域中的潮流:定性方法的多元使用开始流行起来。

基于同样的思路,社会学家Brewer和Hunter(1989)也提出多元方法路径要优于单一方法研究,因为多元方法为资料三角测量法奠定了基础。Brewer和Hunter不喜欢“从研究方法的基本类型中借取一些元素”而组成“合成的”方法(Brewer and Hunter,1989:80)。尽管他们承认合成方法有其作用,但仍认为各种基本方法在掺进对立竞争性的方法论时,就会丧失其本身的某些优势。而且,他们也认为这种方法论折中主义并未为“跨方法的比较”提供足够的资料。在他们二人看来,不同方法的三角测量法可以为因果关系推论提供更好的机遇。

“混合方法设计”一词的意思是什么

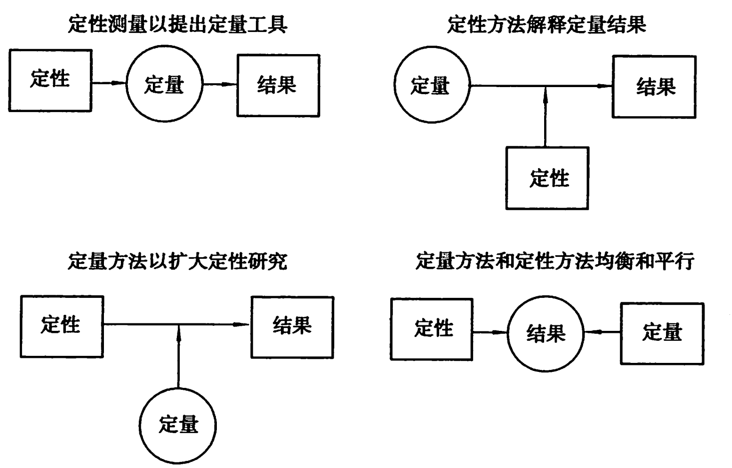
前面我们已经提到,不同领域的作者在最初界定混合方法设计时,都是在三角测量法的名义下进行论述的。Creswell(1995)最近指出,混合方法设计的功能现在已经超出了三角测量法(意指对结果的聚合)。在一篇详细的文献回顾中,Greene等人(1989)界定了使用混合方法设计的五个目的,以及与选择设计方式相关的几个设计因素(见框3.1)。

框 3.1　混合方法研究的五个目的

Greene 等人(1989)考察了 1980 年代以来 57 篇应用混合方法论的论文,列出了这些研究的五个目的:①三角测量,即寻求研究结果之聚合;②补充,即检验某一现象的相同方面和不同方面;③创造,即发现一些悖谬、矛盾或新视角;④推进,即依次使用不同的方法,在第一种方法得出结果后,据此接着使用第二种方法;⑤扩展,即采用混合方法,扩大某一项目的广度和范围。Greene 等人(1989)同时举出了影响特定的混合方法研究类型之选择的设计因素。这些设计因素包括方法的特性、所研究的现象、范式框架、不同方法的相对地位以及实施这些方法的标准。

Greene 等人注意到,依借混合方法研究进行设计选择时极为复杂,这反映出了目前在混合方法路径上所存在着的混乱。第 1 章已经提到,Datta(1994)断言评估者正在使用的是她所说的"杂糅模型",因为混合方法研究并没有一以贯之的范式或理论。包括 Creswell(1995),Greene 等人(1989),以及 Morse(1991)和 Patton(1990)在内,好几位论者都作出了努力,试图建立一个关于混合方法设计的分类法。框 3.2 给出了一个案例,它来自一项有关生育健康问题的跨国研究(Ulin,Waszak and Pfannenschmidt,1996)。

框 3.2　定性方法和定量方法相结合的不同场景



来源:Ulin 等人(1996)。经授权使用。

我们为理清这些不同样式的混合方法设计提出了一个类型学,表 1.1 对此已经作出了介绍。本节的后面将简要地对以下的混合方法设计进行介绍:

- 同等地位设计(Equivalent status design):顺序(定量/定性和定性/定量)设计以及平行/共时(定量 + 定性和定性 + 定量)设计;
- 主次设计(Dominant-less dominant design):顺序(定性为主/定量为次,定量为主/定性为次)设计和平行/共时(定性为主 + 定量为次,定量为主 + 定性为次)设计;
- 多层次路径设计。

需要注意的是,如果说资料收集的类型与研究中所用的资料分析类型纠缠在一起,那么,混合方法一词通常既指涉资料收集的方法,亦指代资料分析的方法。在本章下一节讨论混合模型研究的问题时,我们将分开处理资料收集与资料分析。

同等地位混合方法设计

在同等地位混合方法设计中,研究者将平等地使用定性、定量方法来理解所研究的现象。

Morse(1991)在一本关于护理研究的作品中,声称定性与定量两种范式在一项研究中不可能是地位平等的。然而,我们在教育学研究中的经验却不是如此,在框 3.3 中就可以看到一个同等地位设计类型的案例。

主次混合方法设计

Morse(1991)给出了护理研究领域中数个单独研究的案例。在这些研究中,有一种主导的范式及其方法,但在总体中仍有一小部分研究内容采用了另一种研究设计。这些研究案例被认为是主次混合方法的研究设计(见框 3.4)。

在那些有史以来始终被一种范式所支配的研究领域(如定量方法之于实验型心理学、定性方法论之于人类学)中,主次型的混合方法设计可能会流行开来。例如,在实验社会心理学研究中,往往会采用一种非正式的、通常较松散的"实验后的访谈",以此

来检测有关人员对主要假说的理解意识，并收集对本项研究之主观反应的资料，这些意识和反应有可能会影响到本项研究的结果。对于研究者而言，实验研究的这些组成部分显然属于“附件”，其重要性与假设—演绎所驱动的方法（hypothetico-deductive-driven approach）不可相提并论。

框 3.3 一个同等地位混合方法设计的案例

在对津巴布韦修纳人*婚姻模式的研究中，Meekers（1994）将资料收集与资料分析的定性方法和定量方法结合在一起：定性资料收集将先前的研究者所作的民族志研究包含在内，由此形成了对这个群体内部的联姻类型的界定；定量资料收集则使用了 Udry、Dole 和 Gleiter（1992）所提出的一种事件史研究方法。研究中使用了一块有许多插条的金属板，插条代表着婚姻过程中所发生的事件。受访者被要求拔掉他们所没有经历过的事件之插条，并将剩下的插条按发生顺序排列好，然后一一描述所真实发生过的事件。这些描述被记录在问卷之中。

通过将事件史研究的受访者再归入民族志研究所发现的不同类型，Meekers 将民族志的研究结果与事件史研究的结果融合在一起。这些分类包含了三个方面：联姻的类型、联姻形成中的家族干预、家族对夫妻行为的控制。

联姻的类型通过观察每一类型中的受教育程度、居住地和年龄的比例来进行分析；家族的干预程度由在每一联姻类型中，经历过许多次可被视为家族干预事件的受访者的所占比例来表示；家族控制则由在任何一种明确的家族干预事件之前已发生初次性交的受访者的比例来界定。Meekers 的研究结果印证了民族志研究的发现，它表明年轻的修纳人夫妻已开始摆脱传统的婚姻形式。但她的研究也发现，传统习惯并没有完全被忽视。在联姻之家族控制的问题上，该项研究发现家族对于初次性交和怀孕很难进行控制。但研究也发现，家族对于同居仍然有一定程度的控制。

注：*修纳人是罗得西亚与邻近的莫桑比克的农民——译者注。

顺序型混合方法设计

在顺序型混合方法的研究设计中，研究者先进行定性研究的

步骤,然后再进行定量研究的步骤,或者顺序相反。由于两个阶段差异明显,研究者可以"将每个步骤所隐含的范式假设完全展现出来"(Creswell,1995:177)。Creswell 将此种设计称为"两阶段研究设计"。这种设计在研究生或研究新手中十分流行,因为他们往往希望兼顾这两种方法,但又不愿陷入同时使用二者所带来的麻烦中。在定量—定性的顺序中,研究者先采用定量方法,随后展开定性研究。

框 3.4　一个主次混合方法设计的案例

Creswell(1995)叙述过一个主次设计(定量方法为主 + 定性方法为次)的经典案例。在这一案例中,研究者在检验某一理论的过程中进行了一项实验(定量研究为主),而在资料收集过程中,又进行了一个简短的定性访谈(定性研究为次)。尽管访谈获得的信息十分有用,但在整个理论驱使下的实验设计中仍只是一个小的组成部分。Creswell 总结说,这种方法的优点在于它"提出了一个连贯的范式景象……同时又收集了有限的信息来详细地深掘所研究的某个方面"(Creswell,1995:177)。回到框 2.3 所说的"旁观者介入"实验研究,如果研究者能够就被研究者对"受困的女人"的情感反应进行简短的开放式访谈,那么,这就会是一个"定量研究为主 + 定性研究为次"的案例。

Freeman(见 Freeman and Teddlie,1996)所撰写的一篇论文是顺序型混合方法研究的典型案例。首先,在一个以数据性问卷调查之回答作为因变量的常数比较设计(constant comparative design)中,Freeman 检验了一个关于"进步中"学校校长的回答与"稳定"学校校长的回答之间的差异的预设;随后,其又对被视为"进步中"学校的校长展开了 8 个案例研究,试图确定在他们学校里效率得以不断提高的过程的持续模式。

另一方面,在定性—定量的顺序中,研究者首先对某一相对研究不足的主题开展定性的资料收集和分析,然后根据结果设计出定量研究的步骤。框 3.3 所示的 Meekers 的研究可被视为这种顺序(定量研究被设计为用于确认或复查其他研究者所作的定性研究)的准案例。定性—定量顺序混合方法设计的这一种类型比较常见,因为在大多数定量的问卷调查中,定量的封闭式问题工具往往是在

分析了探索性的定性访谈或对口述资料完成了内容分析之后才提出的。Carey(1993)也提供了一个这种顺序类型的案例(Dressler 所作的一项研究,1991)。鉴于 Carey 对这一研究的评论具有重要的影响,我们在框 3.5 中介绍了他对这一研究的总结。

框 3.5 Carey(1993)对 Dressler(1991)定性/定量方法的总结

Dressler(1991)的著作为我们提供了一个很好的研究案例,说明严格标准化的定性研究方法与定量研究方法是可以结合在一起的。在研究中,他所关心的是经济状况、生活方式、家庭人口因素及来自家庭、朋友和社区的社会援助资源等因素,如何影响到美国南部某社区的非裔美国人群体的消沉情绪,以及它们相互之间的复杂关系。Dressler 说明,他应用开放的民族志访谈着手于资料收集(Dressler,1991:72-73),这样就可以获取有关社区对压力和其他社会状况之认知的信息。在这一步骤所得之结果的导引下,他进一步设计了更具结构化的、定量的问卷调查工具。在做完定量调查后,他又展开了第二轮定性资料收集,从而获取生活史和叙述性的案例研究,以便用于为定量结果提供某种阐释性的场景。这些不同类型的资料使得 Dressler 可以建立多种定量的量表,采用回归相关统计方法,并为研究提供了 8 个详细的描述性案例研究。所有这些方法和不同类型的资料都被用于回答 Dressler 所提出的问题——社会环境是如何影响社区的消沉情绪的。

他将民族志案例研究看成一种同时用来"确定研究资料的统计分析中所观察到的东西,提出新的研究途径,并洞察非裔美国人社区居民中间关于压力之话语"的方法。

来源:Carey(1993)。

当然,依次进行定性资料收集与定量资料收集,或依次使用归纳逻辑与演绎逻辑的过程是反复不断的,可以经过若干次表 2.1 中所描述的循环周期(见 Schuyten,1995;Schuyten and Tashakkori,1995)。同时,还有可能将这两种路径进行另外一些形式的创造性整合。例如,Floyd(1993)将其抽取的样本分为两组:一组采用定性—定量的顺序进行研究,而另一组则采用定量—定性的顺序进行研究。框 3.6 简要介绍了她的研究。可见,两阶段设计是顺序混合方法设计中最为简单的。

框3.6　另一个顺序型混合设计的案例

Floyd(1993)结合使用了定性方法和定量方法以探讨四个问题:①成年人所报告的睡眠焦虑有哪些?②这些焦虑与年龄和性别相关吗?③定性方法和定量方法的使用顺序对研究发现有影响吗?④两种方法所得到的结果是一样还是不一样?Floyd使用配额抽样和目的抽样建立了样本,其中包括来自三个年龄级别的男人与女人。定性资料的收集由半结构化的访谈完成,定量资料收集则由一系列自答(self-report)问卷组成。研究者所设计的一份问卷——"睡眠干扰问卷"提供了两种数值:睡眠焦虑频率的数值以及睡眠焦虑强度的数值。

在各个年龄和性别组内,研究对象被随机分为两半——一半以首先接受定性访谈开始,另一半以首先接受定量问卷调查开始。定性分析包括Leininger(1985)的主题和模式分析的几个步骤,由此得到四个主题;定量分析则是一个包括t-检测、ANOVA和相关性分析的统计程序过程。关于第一个研究问题,研究结果显示,健康老年人的睡眠焦虑包括健康因素;睡眠模式的改变以及环境因素;定性分析显示,与年龄相关的睡眠焦虑是环境因素与睡眠模式的改变,二者都会随着年龄的增长而呈下降趋势;定量分析没有显示出有任何与年龄相关的焦虑因素。两种分析都体现出性别差异:男性和女性所受的干扰来自于睡眠的不同方面。

当调查问卷的填写在访谈前完成时,两种方法的使用顺序对研究的成果产生了影响:访谈中受访对象使用的语言受到了调查问卷的影响。两种方法的某一些结论是类似的,但也有很多不同。这两种路径的结合使用特别适合于有必要进一步研究的问题或领域,而如果只用一种方法,那么就无法弄清楚这些问题或领域。

平行/共时混合方法设计

在平行/共时混合方法设计中,研究者同时采用定性方法和定量方法收集资料,并通过二者互补的方式对资料加以分析。尽管Creswell(1995)认为,定量研究的结果并不一定要在这些设计中与定性研究的结果相联系,也不必确认定性研究的结果(反之亦然),但是,使用这一方法的大多数研究产生了回答相同问题的数据资料和描述资料。譬如,在一个简单的研究案例中,研究者

要求教师完成一份关于学校究竟在多大程度上实现了重组的封闭式问卷调查，同时又对学校的管理人员就同一主题进行了开放模式的访谈记录。当研究者对教师的数据性资料以及管理人员的描述性资料进行分析时，他们就会看到这两类资料所体现的对学校重组程度之认知方面的差异性。框 3.7 里是 Timberlake（1994）所进行的一项关于无家可归儿童之研究中的平行/共时混合模型设计。

框 3.7　一个平行/共时混合设计的案例

Timberlake（1994）在一项关于无家可归儿童的研究中，同时采用了定性路径和定量路径进行资料收集和分析。就概念层面而言，尽管这项研究本质上立基于有关自我概念之于适应能力的作用的既有理论之上（这是演绎逻辑），但它同样是开放的，可以发展出关于自我概念对那些无家可归儿童之影响的新理论（这是归纳逻辑）。研究的一个主要目的，是力图发现"无家可归的情境对每个儿童的个人意义"所充当的角色。

研究者对无家可归的儿童以及他们的母亲进行了访谈，同时要求他们的教师完成一份调查问卷，由此完成资料的收集。对教师的问卷包括一份此前已验证过的清单——"心理功能清单"（Psychological Functioning Inventory），用以测量儿童的适应能力。

研究者同时采用了定性和定量两种方法进行资料分析。定量分析包括对统计指标进行计算，并测量适应良好的儿童与其他儿童之间的显著性水平。定性分析则是根据儿童对无家可归情境的理解，建立一个 3×2 的网格。其中一个维度将"需要关注的实质问题"（substantive preoccupation）分成了三个层次：分离/失散、看管/养育、安全/保护；另一维度则描述了两个层次的行为取向：被动抱怨与恢复。同时，还根据儿童的说法识别出了适应无家可归之情境的三种模式，每一种模式又包含一些亚类型；根据教师的回答则区分出成功的学生与不成功的学生。然后，再根据适应能力指标的数值对二者进行比较研究。

统计比较（*t*-检测）显示出这两种学生群体之间存在着很大的差异。论文最后详细生动地描述了儿童的家庭、学习和个人生活，刻画出"学习上不成功的学生的总体形象"。

多层次路径研究

无论是在定量研究的传统中,还是在定性研究的传统中,多层次研究都很常见。这类研究中的资料取自于多个层次的组织或群体,以便获得关于行为和/或事件的更具综合性的推论。例如,在教育学研究中,在学生层次上所收集的资料会与有关教师之属性、学校之特点的资料联系起来。研究者也并不是将学生层次上的变量"求平均"以获得班级层次上的资料,相反,研究者会使用复杂模型同时分析学生层次的资料和班级层次的资料。这种研究的一个定量案例就是 Bryk 和 Raudenbush(1992)所应用的多层次建模(Hierarchical Linar Modeling,HLM:等级线形建模)。

在多层次研究中,也可以在一个层次上应用定量研究方法收集资料,而在另外的层次上用定性方法收集资料。这类研究的一个案例是,在三所高中学校对 1 400 名学生展开调查,同时又对这三所学校的校长进行详细的访谈,并广泛地观察这三所学校内部的社会-心理动力现象。学校层次上的定性资料可以使学生层次上的资料有更丰富的意义,并且更易于理解,反之亦然。由此可见,这两种路径既可以同时使用,也可以依次使用。

与混合模型研究相关的设计问题

本书的一个主要目的就是为实现定性路径和定量路径的融通而提供一种不同的取向,我们将这一取向称为混合模型研究。这类研究设计在表 1.1 中已作过简单的介绍,我们在下节将提供有关这类研究设计分类的更详细的信息。第 7 章和第 8 章则进一步提供混合模型设计的案例,这些案例在其研究过程的不同阶段中都混合使用了定量研究路径和定性研究路径。

对方法的过度依赖:超越方法层次上的单纯混合

对于方法论问题的关注和思考贯通着整部美国科学史(如 Bannister,1987;House,1994;Ross,1991)。这种方法论的偏执在一定程度上导致了范式之争,但也由此引发了 1960 年代到 1980 年代对混合方法设计的应用,进而促使范式之争画上了休止符。House(1994:19-20)对这种方法论的偏执作出了这样的概括:

> 在早期的发展中，美国的社会科学界由于强大的政治、社会和意识形态压力，回避了某些内容的研究主题。他们将精力集中在方法论问题上，希望以此获得价值中立的、政治无涉的和没有麻烦的研究发现，且能与美国例外论的不变主张保持一致……这种立场带来了一种恶性的科学主义，它固执地将方法作为社会研究的核心所在……对方法的过度强调导致了彼此对立的界定：如果一种方法是定量的，那么另一种方法就是定性的；如果一种方法是客观的，那么另一种方法就是主观的。

当那场在很大程度上围绕着方法问题而展开的范式之争得到解决的时候，这一解决主要是一种方法论意义上的解决也就是不可避免的了。新的混合方法设计再次关注于方法论，但却较少关注其他的研究步骤，如问题的形成以及与源自研究结果的结论相关联的推论过程。尽管情况大体如此，但也有一些作者（如 Brewer and Hunter，1989；Creswell，1995；Patton，1990；Sechrest and Sidani，1995）探讨了在研究过程的多个阶段而不仅仅是方法论整体层次上，整合运用定性研究和定量研究两种范式。在关于他们所谓的"多元方法研究"的论著中，Brewer 和 Hunter 讨论了将多元方法路径运用到研究中的所有步骤，包括问题的形成、理论的建构与检验、抽样、测量、资料收集与分析，以及撰写报告。

与此相似，Creswell（1995）也描述了一种"混合方法论设计"。在这种设计中，研究者在研究过程的多个阶段混合使用定性研究和定量研究两种范式。他还讨论了在撰写论文导言、引证文献和理论、陈述研究目的、说明理论假设或交代研究问题、说明研究方法以及撰写研究结果等各个环节上，这种方法论的混合会如何展开的情形。Creswell（1995：178）总结说：

> 这种方法增加了研究设计的复杂性，但同时也利用了定性研究和定量研究两种范式各自的优点。而且，整个设计也许最佳地体现出了研究者意图在研究过程中往复于归纳式思维和演绎式思维之间的过程。

由此可见，在研究的不同阶段混合使用定性研究和定量研究更准确地反映了图 2.1 所示的在演绎与归纳推理之间循环往复的研究周期。

混合模型研究:混合方法设计的一种替代

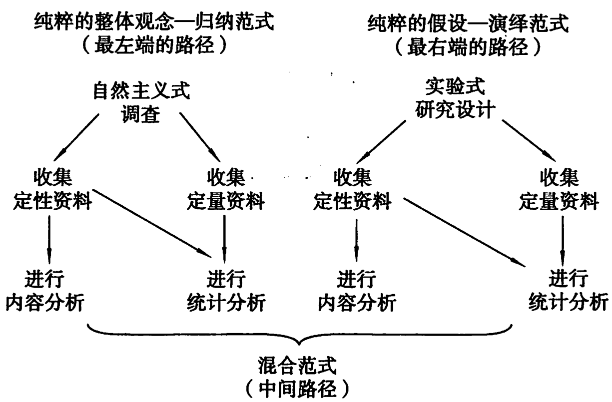
在本书第 1 章曾经提到过,我们使用的混合"模型"研究这一概念,所指称的是这些研究超越了方法论的混合,还包括在研究过程的其他阶段进行方法的混合。仅仅在方法上混合是非常不够的,定性—定量之间的区分也并不仅仅限于"方法"方面。同时,如果仔细阅读社会和行为科学领域最近的许多文献,就会发现研究者们已经在他们所从事研究的多个阶段混合使用了不同的范式,尽管他们所用的设计类型并没有很好的分类法对之概括。

Patton 的"纯粹形式研究"与"混合形式研究"

在对混合模型研究的下述讨论中,我们会在研究过程的三个主要阶段上将定性范式与定量范式的几个方面都混合起来。Patton(1990)在其命名的"混合形式"设计的讨论中,率先提出了在这三个特定阶段中的混合范式。我们在这里对研究过程的三阶段所做的概念化工作就是在 Patton 的原创模型基础上发展而来的。

Patton(1990)在谈到他所说的"方法论混合"时,提出了三个组成部分:①设计(自然型研究/实验型研究);②测量(定性资料/定量资料);③分析(内容分析/统计分析)。将这三个维度加以整合,Patton 获得了 6 种不同的设计(即 2 × 2 × 2),包括纯定性路径、纯定量路径以及另外 4 种"混合形式"路径。框 3.8 描述了 Patton 对这些混合形式所做的概念化工作。

框 3.8 Patton 对纯粹研究和混合形式研究的分类



确定混合模型分类的三个维度

我们之所以青睐 Patton 的这个分类办法，是因为它很实用：第一，你如何设计你的研究及研究问题的框架？第二，你收集什么样的资料？第三，你如何分析并理解这些资料？这三个问题都是我们在方法论课堂上或论文委员会的指导过程中提问最多的问题。

但我们仍对 Patton 的三个维度进行了扩展，并有如下说明：

1. Patton 的设计维度（自然主义式调查/实验式研究设计）可以扩展为“研究类型”的维度（重新定名为探索型研究/确证型研究）。立基于这一区分的另一维度是分辨有的研究有预设（确证型研究），有的研究没有预设（发现型研究）。

我们此前已经提到，就对外生变量和误差幅度的控制而言，自然环境与受控环境的区分是一个连续谱，而非截然分明的二分法。而假设的存在与否是一个非此即彼的二分法，它将现象研究过程中不同阶段的截然分开，强调探索型研究之后才有确证型研究。

这里的第一个维度包括研究问题或假设的形成。在探索型研究中，研究的目的通常以研究问题的形式表述出来；而在确证型研究中，至少存在某个研究假设，即已经对结论进行了预测。

2. Patton 的测量维度（定性资料收集/定量资料收集）可扩展为定性资料的收集与操作，以及定量数据的收集与操作，其涉及的是研究资料的形式。我们认为，对描述资料与数据资料的区分有利于建立一个易于理解的二分法（Miles and Huberman，1994），尽管有些学者认为这一区分过于简化（如 Hammersley，1992）（本书第 6 章将进一步讨论如何将描述资料和数据资料转化为其他形式的资料）。

对这一维度的重新命名是为了实现明晰化：在定性资料与定量资料的收集过程中，会伴随不同的研究操作，包括不同的测量技术、不同的结论信度和效度的建构方法、不同的抽样程序等。本书第 4 章和第 5 章将对这些研究操作展开讨论。

3. Patton 的分析维度(内容分析/统计分析)还可扩展为定性分析及定性推论对统计的分析和推论。这个扩展了的第三个维度在两个方面不同于 Patton 原有的维度:定性分析取代了内容分析,并且加入了推论过程。

定性分析之所以取代内容分析,是因为前者更具包容性,其在诸多具有广泛影响的文献中已经得到了广泛的采用(如 Denzin and Lincoln,1994;Miles and Huberman,1994;Wolcott,1994),而后者通常被认为是传播学领域所采用的术语(如 Berelson,1952)。在这一维度中加入推论是因为推论与分析是交叉在一起的,并且我们也希望这一分类法不会局限于方法论和分析问题之上。如果去阅读一篇已发表的“定量研究”的文章或学位论文的结论部分,那么就会发现,作者所作的总结和阐释远远超越了他们所掌握的结果。每位研究者都可以选择恪守在其发现的界限之内,或者通过推论扩展其结论。

上述三个扩展了的维度代表了 Krathwohl(1993)和 Gall 等人(1996)所描述的研究推理链条的 9 个步骤:

探索型研究对确证型研究的维度可以包括第 1 ~3 步:对以往研究的总结,提出解释、三段法推论、理论或观点,陈述问题、假设、预测、模型。

定性资料的收集和操作对定量资料的收集和操作的维度则包括第 4 ~5 步:设计或结构化研究,收集资料。

定性分析和推论与统计分析和推论的维度可包括第 6 ~9 步:概括资料,确定结果的意义,作出结论,开始下一研究。

尽管 Krathwohl(1993)所划分的步骤是针对定量研究而言的,但在大多数情况下,这些步骤也同样适用于定性研究和混合研究。我们之所以将这三个维度进行扩展,而不仅仅限于研究方法(如设计、测量、分析),是为了在研究过程的其他阶段也讨论定性—定量维度。正是在这些维度基础上所作的分类法,包含了混合模型研究。这不仅超越了单纯对方法论的混合,而且将研究过程的其他阶段也纳入了混合的进程。

这三个维度也可以指研究过程的不同阶段:第 1 阶段,包括研

究者的观点和研究问题(或假设);第2阶段,包括设计问题和收集资料;第3阶段,包括资料分析和总结。

对混合模型研究的分类(研究阶段内的单一运用)

根据上面提到的三个维度或阶段的二分法,我们已提出了一个在逻辑上包容所有可能的关于简单混合模型研究的分类法。这些简单混合模型设计所涉及的是在研究的某一阶段单一地运用定性或定量方法,但它们可以组成更复杂的混合模型研究设计,包括在研究过程的特定阶段多次运用定性和定量方法。

整个分类表立基于上面所提到的三个维度或研究步骤:①研究进程中研究类型的维度或阶段:探索型研究与确证型研究;②研究过程中的资料收集和操作类型的维度或阶段:定性资料收集和操作与定量资料收集和操作;③研究过程中的分析和推论类型的维度或阶段:定性分析和推论与统计分析和推论。

这三个维度的2×2×2交叉分类产生了8种不同类型的研究模式,这与Patton(1990)对研究设计的分类框架大体相似。同样,在这样分类所产生的8种类型中,有两种是传统的定量模式和传统的定性模式。另外6种类型则属于混合模型研究,它们在研究过程的不同阶段混合了定性研究与定量研究。(关于这种类型框架的说明,见表3.1)。

本书第7章和第8章将详细讨论下列6种混合模型研究:

1. 混合模型研究类型Ⅰ、Ⅱ是确证型研究(见表3.1中的类型Ⅰ和类型Ⅱ);

2. 混合模型研究类型Ⅲ、Ⅳ是探索型研究(见表3.1中的类型Ⅲ和类型Ⅳ);

3. 罕见的混合模型研究(类型Ⅴ和类型Ⅵ)由于涉及对定量资料的定性分析,不太常见,但我们在第7章和第8章也将对这两种类型也进行讨论,第6章则描述了“定性化”,即将定量资料转化为定性资料(或将数据转化为文字);

4. 混合模型研究类型Ⅶ、Ⅷ则更为复杂,我们将在本章的最后以及第8章对其进行讨论。

表 3.1 关于研究方法的分类

根据确证型研究/探索型研究、定量操作/定性操作、统计资料分析和推论/定性资料分析和推论这三个维度所作的分类

确证型研究				探索型研究			
定量资料/操作		定性资料/操作		定量资料/操作		定性资料/操作	
统计分析和推论	定性分析和推论	统计分析和推论	定性分析和推论	统计分析和推论	定性分析和推论	统计分析和推论	定性分析和推论
纯定量	混合类型Ⅴ（罕见）	混合类型Ⅰ	混合类型Ⅱ	混合类型Ⅲ	混合类型Ⅵ（罕见）	混合类型Ⅳ	纯定性
第2,7章	第7章	第7章	第7章	第7章	第7章	第7章	第2,7章

注:本书第8章讨论的是更为复杂的混合类型,涉及两个类别:平行混合模型设计(类型Ⅶ)和顺序混合模型设计(类型Ⅷ)。

混合模型研究的两个一般类型

本书第1章指出，混合模型研究可以分成两种一般的类型：同一阶段上的单一方法运用和同一阶段上的多元方法运用。本书第8章则在讨论多元方法的应用中指出，多元方法的应用研究设计可区分为：平行混合模型设计（类型Ⅶ）和顺序混合模型设计（类型Ⅷ）。

这些运用多元方法的研究设计是真正意义上的混合研究，因为在研究的设计、资料收集、资料分析和推论等各个阶段上，都同时包含了定性取向和定量取向。譬如以下所举的这些案例：

- 在计划研究项目时，可以将研究假设（表示这是一项确证型研究）和研究问题（表示这是一项探索型研究）混合起来；
- 在开展一项研究时，可以将定性资料的收集（民族志访谈、无结构的观察）与定量资料来源（结构化访谈草案、观察式定比量表）混合起来；
- 可以用循环交替的方式对定性资料和定量资料进行混合分析和阐释，以便用叙述性结果来拓展数据性结果的意义，或用数据性结果来加深叙述性结果的内涵。

研究的方法与策略

第二部分

在接下来的三章中将涉及抽样和测量的问题、资料收集的策略以及资料分析的策略。在这几章里，我们仅仅提供了屈指可数的几个独特的混合方法或混合模型策略，因为其数量本来就不多。从方法论角度来看，混合方法和混合模型同时借用了定量研究的视角和定性研究的视角。因此，使用混合方法或混合模型的研究者必须同时熟谙定量研究和定性研究这两种传统方法，还必须能自如地在两者间实现转换。在这三章里，我们会以多少有些不同的方式分别处理定量研究路径和定性研究路径的混合。

第 4 章关注的是抽样、资料/信息的收集，以及所收集资料/信息的质量问题。我们将探讨资料/信息的信度与效度问题，并对定性路径和定量路径所收集资料的质量进行评估比较。本章还讨论了在资料/信息基础上所得的结论、推断的效度和可信性等相关联的若干问题。我们在这些设计效度的问题上对定性路径和定量路径进行比较，并试图区分“设计的效度”（即“推论或总结的质量”）与“信息/资料的效度/信度”（即“资料/信息的质量”）。

第 5 章关注的是资料收集的程序。定量研究路径和定性研究路径对文中所列举的五种基本的资料收集方式（自我记录、问卷、观察、互动分析、档案资料/文化遗物（cultural artifacts））都有不同程度的应用。因此，我们在关于这些资料收集策略的章节里提出了一个连续谱，用以说明定量研究和定性研究是如何各自不同地使用这些策略的。我们在这一部分再次强调，使用混合方法

和混合模型的研究者完全可以娴熟自如地交替使用这些资料收集技巧。

第6章关注的是资料分析的技术。我们提出了一个有关传统的定量分析策略和定性分析策略的简化矩阵,同时简要讨论了每种传统内部的独特分析。这一章还包括了一些纯粹属于混合方法或混合模型的资料分析策略。我们举出了三种混合分析策略,即转换资料类型、鉴定/确认分析和类型学划分。这一章所强调的是,研究者要灵活自如地使用两种传统研究路径,并且懂得如何能有效地在混合策略中将这些分析技巧结合起来。

4

抽样、测量及推论质量

本章所讨论的几个不同的方法论主题涉及为特定的研究项目收集资料的问题。在一篇学术期刊的论文中，论文的第二部分通常以“研究方法”为题交代其有关研究方法论的信息；在一篇学位论文中，方法论的情况则往往（但并不总是如此）在第3章中给出说明。我们将在本章中评述这方面的问题，特别是关于抽样和测量的问题，因为二者都是研究者在研究论文或研究项目的“研究方法”部分需要讨论的问题。

“研究方法”部分的内容构成

在研究报告中的“研究方法”一节里，研究者要对如何进行研究提供几种重要的信息。通常情况下，研究者需要以某种形式包含以下4部分内容：

1. **抽样/数据来源**。在这一部分，研究者需要界定其观察/分析的单位（谁研究、研究什么），说明是如何选取这些观察/分析单位的（抽样程序），说明样本的规模（或观察单位的数量），等等。

2. **变量及其测量**。在这一部分，研究者首先要逐一界定若干个变量，并确定变量之间的可能关系（如依赖、独立、预测、标准）。对于那些主要关注点不是因果关联的研究课题（如定性研究、关系研究或描述性研究），诸如自变量或因变量之类的术语，因与主题并不相关则不必使用。然后，研究者须介绍将如何记录、观察和测量这些变量，随之还须说明将如何评估这些信息/资料的质量。我们在本章的后面部分将讨论一些确定资料/信息质量的可用方法。

3. **程序**。在这一部分,研究者需详述其研究方法、设计以及研究的程序。如果此前已经叙述过资料收集和(或)测量程序,那么此处便毋庸赘述,但所有其他的程序仍需进行讨论。就某些定性研究而言,研究设计会随着研究的展开而生成,但对这一生成的演变过程要在有关程序的这一部分作出详细叙述。框 4.1 列举了一个案例,摘自一篇定性取向的研究论文(Johnson and Pajares, 1996);框 4.2 则提供了另外一个与资料分析程序有关的案例(关于资料收集程序的更多细节,可参见第 5 章)。

框 4.1 对"共同决策(Shared Decision Making, SDM)"之定性取向研究的研究程序说明

资料是通过多种多样的途径收集而来的,其中包括定期观察、访谈、文件和录像带。研究小组的成员在课题持续的三年时间里,每年对这所学校进行若干次重访,观察其校董会会议、校委会会议、部门会议、课程表讨论、家长会/教员运作以及各种非正式的集会。辅之以若干次会议的全程录像带的田野笔记,均在事后得到了深入的研究。田野笔记的性质取决于为收集资料而进行访谈或观察时所适宜使用的特定方法:一些田野笔记是在接谈的过程中用手工记录的,另一些田野笔记是在接谈间歇的私人休息时间里用录音记录下来的。SDM 的校董会会议往往由多媒体专家全程录像,而且,正是借助这一方法,我们才能够观察并分析这些我们未能亲自参加的会议。我们试图从过程参与者的角度理解 SDM,因此,我们最大的资料获取途径是访谈。我们对 92 位教职工、学生和家长进行了 167 次现场访谈和电话访谈,其中的 156 次用录音机作了记录。各次访谈的时间从 10 分钟到 1 小时不等,平均耗时 30 分钟。访谈是半结构性的(见 Spradley, 1980),尽管访谈的问题在研究课题的整个进程中有所变化,但我们始终鼓励参与者从他们个人的视角谈论学校的 SDM 过程。所有的访谈都登录在本文的附录之中,田野笔记还加入了附记,提供了访谈录音中不可能捕捉到的有关访谈场景的相关信息。我们还收集并分析了与 SDM 过程相关的文件资料,包括所有校董会会议记录的复印件、SDM 程序的规则和指导方针、校委会的报告、向学区和州政府所提交的放权要求,我们还获得并审读了学区就 SDM 事项所发行的各种手册。

来源:Johnsont 和 Pajares(1996:605-606)。

框 4.2　关于资料收集和分析程序之说明的案例：来自一份定性研究取向的研究报告

资料收集程序

本项研究的资料收集包括两个阶段：学生集体会面和学生个体会面。

学生集体会面。第一阶段由两次小组会议构成：这个研究室里的拉美裔学生以及另一个更大的研究室里的拉美裔学生在这两次小组会议期间与主要研究者会面接触。这两次小组会议都是以西班牙语进行的。会议期间，学生们听取了研究目的介绍，填写了背景问卷，并完成了对其已有知识的摸底测验。第二次小组会议的目的是为学生们提供机会，使他们能够实践自言自语的程序。因此，学生们在这一阶段观看了两盘录像带，第一盘放映的是一个只会西班牙语的孩子在阅读时自言自语，第二盘放映的是一个只会英语的孩子在阅读时自言自语。在对录像带讨论后，学生与其伙伴结对尝试进行自言自语。研究者多次鼓励学生想一想他们在阅读时的行为，并回想双语能力对他们的阅读有何影响。主要的研究者还另外会见了英裔学生，并重复操作了对拉美裔学生所使用的相同程序，但所用的语言是英语。

学生个体会面。资料收集的第二个阶段由个别会面构成，期间每位学生都与主要研究者会面交谈。就 Catalina 和 Pamela 而言，共进行了 3 次会面，但对不能说西班牙语的 Michelle 而言，只有 2 次会面。在这些会面中，学生们实践了经提升的自言自语和未经提升的自言自语。在读完每篇文章后，他们都被要求复述这篇文章……研究者还在这些个别会面期间对学生进行了访谈。访谈中，研究者鼓励有双语言能力的学生使用他们觉得最自在的语言。当然，Michelle 是使用英语进行这些访谈的。资料收集的程序最终产生了关于每个拉美裔学生大约 220 分钟的资料，以及关于英裔学生 120 分钟的访谈、回忆和自言自语的资料。

分析

研究者将来自对已有知识的评估、访谈草案和自言自语实践的各种资料汇集起来，建立起了三类学生的个人肖像。研究者运用了包括编码、回忆记录、主题描述和表达等在内的定性研究程序……在肖像形成以前，研究者首先识别出了学生在自言自语实践中所使用的阅读策略及其在访谈中所提到的阅读策略……三名研究者对这三位学生以及其他学生的评分进行了再三解读，以期建立起关于阅读策略（如呈现策略）的生成性框架，将与这一框架相符或不符的正负两类案例都考虑进来……最后，对阅读策略给出了界定：阅读策略是阅读者为了理解他（或她）正接触的印刷材料而作出的公开有意的努力或行动。

来源：摘自 Jimenez，Garcia 和 Pearson（1995）。1995 年版版权归属美国教育研究协会。摘录得到了出版者的允许。

4. **资料分析技巧**。在这一部分,研究者要提供一个概述,说明要如何分析研究所获得的信息/资料。框 4.2 列出了一份定性研究报告的资料收集和分析的程序说明(Jimenez, Garcia and Pearson,1995)。读者在读完这些部分后,会对研究者如何收集、分析资料形成比较清晰的印象。

与上面的案例不同,在研究计划书和学位论文中,有关方法说明的文字部分往往是长篇大论且论说详尽。其中还包含了研究者在完成资料收集工作之后,将对这些资料进行深入分析的第一个计划步骤。就期刊论文而言,由于这些分析的主要内容将会在论文的下一章节(即"研究结果"或"结论与讨论"①)里加以说明,因而资料分析计划的部分内容往往是简练概括的。本书第 6 章将对这些不同的资料分析策略进行更为详细的讨论。

抽样及其与通则化/推论之质量的关系

抽样与结论的通则化特性(外部效度、可转换性)

一般而言,无论在定性研究还是在定量研究中,我们都不能对某个个体集合的全部(即"目标总体")展开研究。即使这一总体的数量较小,通常也只有其中一小部分成为某个研究项目可获得的研究对象(即"可得总体")。因此,如果研究者有意作出能够适用于其他群组或全部总体的推论和(或)政策决定,那么,研究者就应当去研究那些最能代表其"总体"的个体、事件/环境的群组。从这些具有代表性的个体或事件/环境的样本中所获得的结果,更有可能推广适用到整个(可得)总体上。个体、事件/环境的样本所具有的代表性越高,则研究者的研究发现具有"对总体的外部效度"的概率也就越高(关于抽样与结论的外部效度/内部效度之间的关系,见 Henry,1990)。

从定量研究的视角来看,研究发现的外部效度并未局限在相对

---

① 我们支持在"结论"部分前加入"结果与讨论"的综合性部分。有一个综合性的"结果与讨论"的部分就能提供一个更加有动感的方式来提出材料,审问材料的每一方面(如其他可能的阐释、对效度的威胁),并对这些问题进行新的分析。如果在一部分中提出结果,然后再在另一独立部分中对结果进行讨论,就会缺少动感。

于个体之总体的通则化特性(generalizability),还包括相对于研究者所研究之情境以外的情境的通则化特性("生态效度",见 Gall, et al. ,1996),以及相对于研究者所用之概念以外的各概念定义的通则化特性。而且,在测量某个变量时,研究者会抽取并使用一组代表性的事项:许多观察点,以及测量或记录其属性或事件的特定方式。如果将测量、记录这些属性或事件的各种可能的方式视为一个总体,那么,这些已选取的事项、观察点或测量方式实际上就都是总体的一个"样本"。因此,外部效度也就与研究者所得到的结论能否推广到测量概念的其他方式之上的程度有关。由此,研究者(Rosnow and Rosenthal,1996:158-159)大体上可以对"外部效度"给出这样的界定:"这是一种我们可以用来推论这种……关系能否推广到其他类型的人物、环境、时间及方法之上的近似有效性。"

对于大多数定性研究者而言,他们并不想要将研究结论推广应用于其他的个体、环境和时间之上。因为在其看来,操作性假说是受时间和场景限制的(见表 2.1)。而研究结论和推论的通则化特性对于这些定性取向的研究者来说则是不同的,至少在表面上看是全然不同的。一些人将通则化特性说成是"研究结果的可转换性"(transferability),另一些人则认为二者并不相关。根据 Lincoln 和 Guba(1985:297)的说法,"可转换性"的意思是:

> 自然主义者的假设相反:往最好说,只有操作性假说可以抽象化,而它的可转换性则是一个实证问题,取决于发送场景与接受场景之间的相似程度……也就是说,为了(在某一置信区内)确证研究者的推论,研究者需要同时了解发送场景与接收场景。由此,我们从通则化的问题转到了可转换性的问题。如果研究者只了解发送场景的话,就不能作出可转换性的推论。

尽管存在着这些分歧,但我们仍相信,结论/推论之通则化特性的程度(无论冠之以"外部效度",还是"可转换性")①,对

① 外部效度关涉的是分析之中或之后所提出的结论。这些结论和推论是就变量、时间之间的关系作出的。尽管信息、资料的质量会影响到这些结论的外部效度,但二者应当区别开来。应当在对这些变量的关系作出推论/结论之前,对这些资料/信息的质量(效度和信度)作出评估。

于所有的研究者而言都是重要的。我们这些研究者会需要将一个样本的通则推广到一个总体,或从一种情境推广到另一种情境,或将结论/推论从一种场景(特定的环境)转换到另一种场景。

当研究者使用混合方法来考察一个研究问题时,往往会在不同的通则化模式之间进行变换。有的时候,研究者对通则化为一个理论总体更感兴趣;而有的时候,研究者想要将结果/结论转换/通则化到另一种特定的场景之中。这些通则化关注点的差异与本书第2章所描述的科学方法论之轮中的不同逻辑类型有关(见图2.1)。

1. 在归纳模式中,我们从具体的案例中收集资料以建立一个抽象命题或理论:通则化是随着扎根理论的发展从最初的一种特定的环境推广到另一种环境(即可转换性),或者是从一种对思维物进行界定和概念化工作的方式推广到另一种方式(也就是,从特殊和具体到一般和抽象)。

2. 在演绎模式中,我们是在对一个理论总体中假定会发生的结果作出预测:通则化就是从我们所使用的一个特定样本推广到理论总体之中(相对于总体的外部效度)。或者,我们可以从有关思维物的一种操作方法推广到到另一种操作方法,从思维物的一种测量方法推广到另一种测量方法,从一种场景和情境推广到另一种场景和情境。

Babbie(1992)对田野研究(社会学家通常用以指称定性研究和民族志研究的用词)中的抽样以及内容分析的讨论,可以帮助研究者理解抽样与发现/结论的通则化特性/可转换性之间的关系。其(Babbie,1992:292)讨论并总结:

> 田野研究者试图观察其研究现场内的一切事物,所以在某种意义上讲,他们根本不进行抽样。然而,事实上对所有事物都进行观察是不可能的。田野研究者充其量只能观察所发生事物的一部分,这样,他们所观察到的东西事实上就是所有可能观察到的总体的一个样本。如果在一次宗教复兴会上一些人高喊着支持演讲者,那么研究者所听到并听懂的那些喊叫所代表的只是所有那些喊叫的一个样本。或者

> 说，如果研究者在一次骚乱中观察到了暴力行为，那么被观察到的行为只是所有这些暴力行为中的一个样本。研究者很少能够从这些观察中选取控制性的样本，但是他们应当记住代表性这条通则，并相应地解读研究者所观察到的东西。(Babbie，1992：292)

> 在对成文文字的内容分析中，抽样可在以下任何层次上进行：语词、词组、句子、段落、小节、章节、书目、作者，或者与语词相关的背景。其他形式的表达也可以在任何一个适合它们的概念层次上进行抽样。任何传统的抽样工具……都可以运用于内容分析。(Babbie，1992：317)

结论的内部效度和信度

在传统上讲，内部效度的概念被界定为我们可以信任研究者关于变量和事件的因果关系所作结论/推论的程度。根据这一界定，如果研究者相信一个结果变量（"结果"）的变化能归结于一个事先变量（"原因"），而不是其他潜在的发生因素，那么，研究者的结论就具有了内部效度(Cook and Campbell，1979)。正如我们上面所讨论的，我们对内部效度所作的概念化工作并未局限于实验研究与因果关系。我们与一些作者所见略同（如 Krathwohl，1993：271；McMillan and Schumacher，1997：183），认为对非实验型研究及非因果推论（如关系研究和预测研究，其他一些描述性研究以及定性/田野研究，等等）内部效度的问题也是十分重要的。用 Krathwohl(1993：271)的话说，"换个角度来看，内部效度就是一项研究对于将其内含的某些变量联结为某种关系的推论所给予的支持力度"（原文是斜体）。

上述问题尤其适合于对共变变量（即当某一变量高时，则另一变量也高，等等）之间的关系所作的推论。Cook 和 Campbell (1979)将这种形式的内部效度限制为他们所指称的"统计结论效度"的范围之内。用他们的话来说：

> 为什么我们会得出关于共变变量关系的错误结论？想想这个问题的原因是有益的。我们应当说，这些原因（是对正确进行推论的威胁）威胁到了统计结论效度，因为有关共

变关系的结论是依据统计证据而作出的。(Campbell,1979:37,原文为斜体)

McMillan(1996:251)则提供了另外一种有关内部效度问题的概念解释:

> 内部效度一向是作为一个与实验研究中的因果因素相关联的概念而为人所知的。而在定性研究中,内部效度则是一个更宽泛的概念,指的是研究者的分类、解释与真实存在的事实相互吻合的程度:研究者的意思、分类和解释是否反映了现实?资料收集或情境的局限性或实际样式是否扭曲了研究发现?

根据这一更宽泛的内部效度定义,如果研究者自信其所得出的变量之间的关系(因果关系或其他关系)是真实的而非虚假的(不是其他变量引起的结果),那么,研究者关于变量或事件之间关系所作的结论就具有了内在效度。例如,如果说研究者发现男性司机造成的交通事故多于女性司机造成的交通事故,那么,尽管仍不能将性别看成是交通事故的原因,但研究者的结论与内部效度的问题具有了高度的相关性。但是,在性别与事故数量之间所观察到的关系也许仅仅是出于个体的样本、源自对"事故"的界定以及其他原因。如果存在着这些替代性的解释,那么,研究者的结论就不具有内部效度。①

这也就是说,这种自信度直接源自于研究者排除其他关于既知关系的替代性解释,(面对着其他专家、读者等)"捍卫"自己结论的能力。我们当然希望关于这种"捍卫"的一个主要假设能够得到证实,即对一种关系或一个结果存在着一个主要的合理解释,而不同的专家对于这个解释可以达成一定程度的共识。在大多数定量研究中,"捍卫"取决于研究者实际上或在统计学意义上对外生变量的控制程度(即第3章所讨论的最大最小控制原则中

① 另一个例子是在"预测研究"中推断死亡年龄与性别之间存在关系。其结论是:性别不是死亡的原因。如果这一结论是内部有效的(如果这一关系不是出于偶然的结果、拙劣的测量以及不恰当确定的总体等),那么一家人寿保险公司就能够用性别来"预测"死亡年龄。倘若你处在人寿保险公司的决策职位上,你会在你不能确信二者的关系是真可信赖(内部有效)的情况下使用性别来预测死亡年龄吗?

的控制原则);在定性研究中,研究者对其研究结论的确认则要依靠多种方法的分析来实现。例如,McMillan(1996:252)引用Miller、Leinhardt和Zigmond(1988:271)的话说:“要使用多种资料来源来比较研究正在形成的趋势是否能保持连贯一致……只要有可能,在一种资料来源中所辨识出来的趋势就要使用其他资料来源加以印证……最后,还要全面搜寻资料以发现是否有否证的证据。”

只有经过研究者对推论小心而谨慎的检验,内部效度才能加以确定。Krathwohl(1993:271)将这一评估过程概括为5个判断:

> 首先,是解释的信用度和转达的忠实度,它们构成了将研究中的变量联结起来的概念依据;接下来的两个判断,是对结果的证明和对替代性解释的排除,它们构成了将变量联结起来的经验依据;最后,需要判断是否存在可信的结论。这最后一个判断所考虑的是,与已有研究的一致程度,以及对概念依据和经验依据之效力的判断。

对于上述五个方面的判断,我们愿意再增加一个“推论一致性审查”。这是指推论和阐释要与所获得的数据/信息分析,以及同一研究中其他推论和结论保持一致的程度。下文就是对这六个判断/活动所作的简单描述(前述五个方面的详细内容,可见Krathwohl,1993:271-280):

1. 解释的信用度(credibility),是指对变量间关系的解释在理论上和概念上的合理程度和可接受程度。

2. 转达的忠实度(fidelity),是指研究的概念框架(包括问题、假设等)被转达成各种设计要素(即恰当的抽样、测量/观察,以及其他程序)的程度。我们在第2章谈到此时称之为“问题主宰”。

3. 得到显示的结果,这涉及是否发生了某些结果,这些已发生的结果是否合乎预期的判断。

4. 被排除的对立性解释,这意味着要评估针对既有关系有无其他合理解释的程度。在本书第3章,我们在谈及Kerlinger的最大最小控制原则中的控制原则时,对此有所讨论。

5. 可信的结论,这涉及对现有研究结论与文献所载之既有发

现相一致的程度进行评估。在大多数情况下,理论上未预期的结果会使人们在接受其相关推论时变得小心(怀疑其内部效度),但是,如果推论/结论是依循研究发现而作出的,那么,尽管结论出乎预料,合乎上述标准且令人信服的研究也是具有内在效度的。例如,方法论上合理完善的否证性重复研究往往为一些专业期刊所看重。《社会心理学的代表性研究》就完全属于这类研究。

6. 推论一致性审查,这意味着要确定有关推论和阐释与对所获资料/信息的分析相互一致的程度,还要确定在同一项研究中不同的推论和结论之间互不矛盾的程度。这与 Lincoln 和 Guba(1985)所提到的定性研究中"可确证性审查"非常相似(本章后面将对此加以讨论)。

定性研究者在内部效度问题上与定量研究者一般都会表现出不同的倾向,至少在表面上如此。定性研究者更喜欢讨论结论的"可信度"问题,而一些定量研究者则使用"可信度"来讨论内部效度问题(具体的案例见 Henry,1990:15)。同时,正如我们上面所讨论的那样,"可信度"被看作是内部效度的标准之一。定性研究者与定量研究者的主要分歧在于,他们对由谁来充当对"可信度"的"评估者"有不同的看法。定量研究者把对结果可信度的评估交给其他人(如其他专家)。然而,大多数定性研究者对结论的可信度进行评估的途径,却是要确认这些结论对于那些其多元现实被重构和描述的个体是否可信(Lincoln and Guba,1985)。换言之,通过确认他们作为人类资料的收集工具在多大程度上再现了他们的信息提供者所给予他们的多元建构之现实的方式,定性研究者可以确定其结论的可信度(Lincoln and Guba,1985;Spradley,1979)。但这些多元现实相对于他者(包括专家或研究结论的使用者)而言的"可辩护性",并不是内部效度之概念化的组成部分。

研究者可以发现,与通则化特性一样,内部效度与结论/推论的可信度之间存在着相似之处(为使研究者的工作轻松一些,我们希望有一天这两个概念可以整合成一个术语)。我们在上面已经论及,这两个概念之间的一个主要差别在于对评判结论之"可信度"的主体/人物有不同的主张。必须注意的是,尽管被研究的个体也许掌握着关于行为和事件的信息,但对他们而言,变量之

间和事件之间的联系并非总是清晰的(这实际上也是许多精神治疗、婚姻治疗、个体纷争调解和机构纷争化解的一个主要假设)。我们已经说过,内部效度是指源自于联结观察/资料之关系的推论/结论的可信度。因此,将资料来源看成是唯一的、甚至是最好的可信度评估者,并不是没有问题的。而且,许多的定性研究都涉及内容分析(本书第6章将展开讨论),而在这些内容分析中,定性研究与定量研究对于推论质量(内部效度)的看法却只有很小的分歧。最后,即使这两种研究路径之间的分歧确实是重要的,但这种分歧对于混合模型研究者而言也并不是大问题,因为混合模型研究者会妥善地根据一方或双方的标准,灵活地对结论的可信度进行评估。

抽样与结论的内部效度

对个体或情境的抽样("选择偏好")可能会影响到研究成果的内部效度。例如,选择同一总体中的某一些个体而非另一些个体来进行观察,也许就会影响到两个变量间的关系(如焦点小组或调查中的志愿参与者)(关于调查研究的案例,见 Mangione,1995,第5章)。在这些观察结果的基础上所作出的一些推论,其可信度就有可能很低,甚至完全没有。在实验型研究中,如果组群之间在与自变量相关的若干方面有先在差别的话(例如在准实验型研究设计中),就会发生这种情形。在非实验方法的研究中,也不同程度地存在着这种情况,即已经推论出的两个或更多变量之间的关系,其实是研究所用之样本的某些属性所导致的。

这些样本的属性往往是抽样偏好所造成的。例如,如果一项研究对学校系统中的教师进行寄送式问卷调查,但只有学校中过得不开心的教师寄回了问卷,那么,研究者根据这些资料而得到的结论也就陷入了上面所提到的尴尬境地。因为如果根据这样的资料,研究者会给出结论说,在教师的职业满意度上并没有性别差异。但由于在回答问卷的教师样本中,无论其性别如何,他们都普遍过得不如意,这就会直接导致研究者对性别差异的失察(关于邮寄问卷调查若干问题的讨论,见 Mangione,1995,第六章)。同时,"统计结论效度"很低也可能会导致对差异的失察(见

Cook and Campbell,1979:39-50;Henry,1990:13)。

抽样偏好与抽样误差

现在应该已经清楚,我们对于从样本到总体的通则化推广的可信度,在很大程度上依赖于所抽取出的那些个体或事件/情境的"代表性"程度。在考虑抽取样本之前,我们首先需要区分"抽样误差"与"抽样偏好"。抽样误差是随机的,当研究者每次抽取一个个体、文本、情境或任何一个"观察单位"时,这一观察单位都会不同于其所在的总体。因此,研究者总是会在将通则化扩大到观察单位之总体的问题上出现误差(我们只能希望误差不大)。

通常,当研究者随机抽取一个较大数量的观察单位进行观察/研究时,如果其中一些单位在某个方向上偏离了"单位的总体"(例如一些人在经济上处于弱势地位),那么,这种偏离会被那些在其他方向上偏离总体的单位所抵消。例如,如果研究者的样本中既有一些受过高层次教育的个体,同时也有一些只受过低层次教育的个体,那么,样本组群的整体"图像"将会是相对平衡的。

而为了确保随机抽取的单位能够呈现出这种平衡,研究者就必须在其观察/研究对象中抽取出较大数量的样本。那么,研究者"随机"抽取5所高中就可以设想样本中的学生代表了"平均"总体(即不同的抽样误差彼此抵消)吗?研究者在随机地抽取9个学校并对其中的所有教师进行了访谈以后,就可以假定由于学校是随机选择的,因而其便拥有了有关教师群体的代表样本吗?最后,研究者观察一国某地的3个急症室,就可以推论出全国所有急症室的情况吗?但愿研究者现在已经明白,对上述这些问题的答案都是否定的。为使抽样误差接近于零,研究者需要抽取一个较大规模的样本。如果无法抽取较大规模的样本来进行观察/研究,那么研究者就应当根据研究对象的特征来抽取样本单位,而不是随机地抽取。只有这样,研究者的研究结论才能更加可靠且可信。

研究者在研究课题中所观察的数量会同时影响到研究发现的内部效度和外部效度。外部效度通常可以通过增加代表性观

察单位或案例的数量来加以提高。但此处必须强调“代表性”：如果增加的个体或观察单位并不是对研究者之研究具有重要意义的情境中的组群之“典型”，通则化的水平就不会得到提高。在定量研究中，除了提高通则化之水平的需要以外，样本规模还会影响误差边际以及用于检测结果的统计测试之效力，进而影响“研究成果的统计结论效度”。我们将在第 6 章有关资料分析的论说中进一步讨论这个问题。关于样本规模对结论效度的影响，还可以参见 Henry 的著述（1990：13）；关于如何确定样本的规模，可参见 Henry（1990，第 7 章）和 McNamara 的论著（1994）。

与抽样误差不同，“抽样偏好”是系统化的（非随机的）。例如，为了完成一项焦点团体研究，如果研究者从某个星期五下午碰巧在图书馆的学生中，每次“随机”在 5 位学生中抽取出 1 位，那么，研究者所得到的也许就是一个带有偏好性的样本，它无法代表“平均”的大学生的观点。之所以如此，是因为星期五下午呆在图书馆中的学生很可能会与在校学生总体中的其他部分同学完全不同。与抽样误差的不同之处还有，扩大样本的规模并不能降低样本的偏好程度。显然，从一个带有偏好的样本中所得到的结果，对于总体而言并不具有代表性（即研究发现的可转换性或外部效度很低）。

随机样本与非随机样本

有时，抽样单位是由个体构成的完整群体（即“整群抽样”，cluster sampling），如学校、诊所，研究者因而也不可能抽取足够大数量的这类单位或群组以减少抽样误差。那么，在这样的情况下，研究者应该怎么办呢？也许研究者此时更明智的选择是在关于这类单位的已有信息的基础上，进行非随机抽样（进一步的讨论请见 Maxwell，1996：71）。我们将在本章的后面讨论某些“有目的”抽样的策略（关于“非概率样本的功用”utility of nonprobability samples，参见 Henry，1990：23-25）。例如，如果研究者样本学校的民族/种族组成对于所研究的问题具有重要的意义，而研究者却只能抽取 3 个学校进行研究，那么，为了保证研究者的样本具有合乎意愿的民族/种族多元性，研究者可以这样完成对学校样本的抽取：一个少数族裔学生较多的学校，一个少数族裔学生较少的

学校以及一个少数族裔学生人数适中的学校。

这种类型的抽样尤其适用于小规模的深度研究项目。例如,这种方法广泛地应用于那些观察/研究小数量单位的定性取向的研究中。在本书第3章,我们讨论过一项由 Freeman 和 Teddlie(1996)所完成的、出自学校效率问题研究的文献,其中详细说明了这类小规模的研究。在类似的研究中,研究者根据可获得的关于观察/研究单位的定性或定量信息展开对样本的选择。例如,根据对已有的档案资料进行初步的回归分析来选择用于研究的学校样本,并在外层设计中将学校分类成有效率的学校和无效率的学校,这就是利用这类信息的一个具体案例(如 Patton,1990;Stringfield,1994)。我们将在第6章讨论这些初步的预先抽样的资料分析技术。

样本的类型和抽取程序

表4.1和表4.2对概率性抽样程序和目的性/非概率性的抽样程序进行了总结。尽管一些抽样程序是与定性路径或定量路径相联系的,然而大多数抽样程序可以同时用于定性与定量两种研究类型(关于定性研究的概率性抽样的案例,见 Miles and Huberman,1994)。表4.2中所讨论的目的性抽样策略尤其如此。这些策略既广泛运用于描述性的定量研究之中,也广泛运用于定性研究之中。

表4.1 研究中常用的概率抽样策略

抽样类型	特征和抽样程序
简单随机抽样	总体中的每个个体有平等的和独立的机会被选中。样本根据一个随机数表或电脑产生的随机数字,通过偶然性的抽取产生。
系统随机抽样	根据样本所需的数量 n,在目标总体中每隔 n 名抽取一人。这一方法只有在拥有关于总体的随机顺序名单时才能使用。
分层随机抽样	这一方法是在研究者知道总体中的次级群体(层)的比例时才能使用。随机性选取,但从每个分层中选取。
比例抽样	每个次级群体在样本中的比例与在总体中的比例相同。

续表

抽样类型	特征和抽样程序
非比例抽样	不考虑在总体中的比例,样本从每个次级群体中抽取同等数量的个体。研究结果可推广至次级群体而非整个总体。这种抽样策略适用于这样的总体:其中包含一些缺乏足够比例的少数群体,如果用简单随机抽样就无法使这些群体得到体现。
整群随机抽样	以总体中已经形成的个体集群作为抽样的单位。因为集群是抽样的单位,所以必须选择较多的集群,随机抽取5所学校所构成的样本并不是一个真正的集群随机抽样。
多阶段集群抽样	这一方法融合了整群随机抽样与其他抽样方法。例如,首先选择集群(如校区),然后在每个集群中随机地抽取个体/学校,或依据相似于分层抽样某些特性的方法抽取样本。

注:这些抽样策略在定性研究中和定量研究中都可采用。关于概率抽样方法的进一步描述,见 Henry(1990)。

表4.2 研究中常用的非概率抽样或目的抽样的策略

抽样类型	特征和抽样程序
目的抽样	对个体/组群的抽取并非随机的,而是根据特定的研究问题或研究目的以及关于这些个体/组群的已有信息而进行的。
同质性抽样	个案的中选是因为它们具有同样的性质和(或)同等的属性。不考虑样本的规模大小,而只根据研究的目的进行选择,中选个案可能是极端个案或异常个案/边缘个案,具有很高程度的某种属性,或在某一属性上处于平均的或典型的状况。
异质性抽样	个案的中选是因为它们的集合在某种属性(如民族、教育)上提供了最大的异质性,而这种属性对课题的研究目标具有重要的意义。通常会从属性的每个层次中至少选取一个个案。

续表

抽样类型	特征和抽样程序
分层非随机抽样	这一方法与分层抽样相似,但是以非随机的、便利的方式进行抽样的。一个或多个个案是从研究总体的每个次级群体中非随机地(自愿参与者、现成参与者等)抽取出来的。在社会学研究中,这一方法即所谓的"配额抽样"(见 Ary, Jacobs and Razavieh,1996)。
滚雪球式抽样或链式抽样	根据从已抽取的样本成员或其他个体中所获得的信息,进一步抽取个体样本。由于每个个体都可能提供不少于一个的、关于其他合适个案的信息,样本数量因而随着研究的推进而不断扩大。
顺序抽样	最初抽取出一个小的样本,然后继续抽样,直至达到满意的程度为止(见 Krathwohl,1993)。例如,在焦点小组中,当新的焦点小组相对于此前的各个焦点小组已提出的观点和主题不再增添新的东西时(即处于饱和状态),抽样过程就可以终止了。
便利抽样	抽样过程并非立基于相对研究目的或研究问题而言的适切性,而是根据资料收集的可获得性和容易程度而展开的。由此而抽取的样本还包括所谓的"俘获样本"(即调查者能够接触的个体组群,如教室中的学生)、"自愿样本"和"偶然样本"等。

注:这些抽样策略都会应用于定性研究和定量研究中。有关在定量研究和定性研究中所使用的这些方法以及其他一些非概率抽样方法的进一步描述,分别见 Henry (1990)、Patton(1990)。

甚至在定性或定量的传统路径内部,研究者在讨论抽样方法时往往认为概率抽样方法与非概率抽样方法是互相排斥的,每种类型的研究只能使用一种抽样技术。但事实上,社会和行为科学中的许多研究课题并非如此。研究者通常会综合使用那些最适于研究者课题需求的抽样技术。"多阶段整群抽样"就是一个典型,通过这种技术,整个单位(整群,如某校区)被随机抽出,然后在每一个单位的内部,再随机地抽取出更小的单位(更多的细节请见 Henry,1990:30-32)。研究设计甚至会需要研究者将其样本

分布在其中的一个层次，结果就是结合使用了两种类型的抽样技术（如整群抽样和分层抽样）。

与多阶段抽样相似，将概率抽样策略与目的抽样策略混合起来也是可能的（对于调查研究中这种混合过程的复杂性问题，简短的相关讨论可见 Fowler，1993：24-26）。例如，研究者可以用目的抽样方法抽取出 3 所学校，并在每一学校中随机抽取出学生而进行访谈或其他测量。正如我们已经讨论过的那样，单一的样本类型或抽样程序不可能适用于所有的研究问题或研究目的。我们相信，对于抽样策略的区分应当立基于研究问题的性质，而不是根据定性—定量路径的不同进行划分。显然，作为一个混合研究的主张者，研究者可以选择将多元的抽样技术混合起来。

测量的质量和推论的质量

传统的定量研究者所关注的主要问题之一，就是确保资料以有效和可靠的方式收集起来。表面上看，这些问题对于传统的定性研究者并不那么重要，因为定性研究者一般并不在展开研究之前设计精细的工具以测量其变量和主题。然而，定性研究者的确很注重支撑其结论/推论的基础性信息的质量。

与定量研究者不同，定性研究者将信息/资料质量的问题（也可称可信性和可靠性）与设计质量的问题（即前面提及的外部效度和内部效度）结合起来进行考虑。一些定性研究者将设计效度的问题重新界定为可转换性和可信度。正如我们前面所述，从定量研究的视角来看，测量效度是研究过程中所收集的资料/信息的属性，而内部效度和外部效度则是对所有资料/信息进行处理/分析后所得出的结论/推论的属性。因此，内部效度/外部效度是最终研究结果的属性，而测量效度/信度则是达致最终研究结果的过程（方法）的属性。

定性研究者对资料/信息的质量与推论/结论的质量并未作出如此明晰的区分。定性研究传统对这两个组成部分不加区分的一个主要原因，也许是在定性研究中，资料收集和资料分析是密切交结在一起的。对比考察定量研究程序与定性研究程序自会澄清这种差异：定量研究者收集资料、评估资料的质量（效度和

信度),然后进入资料分析,之后其作出结论,并根据推论的质量(内部效度)和通则化水平对结论进行评估;另一方面,定性研究者则在收集信息过程之中或之后不久便会进行他们的"资料分析",然后作出推论/结论,再收集更多的信息,进行"分析",作出进一步的推论和处理。尽管在这种研究中,很难对资料的质量和推论的质量分别进行评估,但我们也认为,对二者的评估标准不必雷同。因此,我们应当尽可能地分别对这两个方面进行评估。

在本章剩下的部分,我们试图从三个部分对信息/资料质量和推论/结论质量的问题进行讨论。首先,我们从总体上讨论与测量相关的问题,特别是在传统的定量研究视角下与测量相关的问题;其次,我们从总体上提出推论质量(内部效度/外部效度)的问题,并特别参照了定量研究者的观点;最后,我们提出了定性研究者对研究的这两个组成部分的概念界定。我们将尽可能地在表述中区分开信息的质量和推论的质量。

信息/资料的质量:定量研究的视角

在一般意义上,测量是一个将标签或数值赋予事件或属性的不同层次、量级或性质的过程。定量测量往往通过赋予不同的数值来区分某一变量的不同量级,定性测量则赋予标签和/或描述性指标来确认事件、人物、情境和行为的不同组群。通过对某一事件的不同量值赋予定性标签或定量级别(数值),我们就可以区分那些不同的量值和层次,并将之记录下来以便随后进行处理。

事实上,围绕"研究方法"的大量口舌之争是围绕着测量方法而展开的。研究者对于定性和定量这两种一般方法的厚此薄彼或褒贬不一的执拗信念,是导致这一争论的部分原因。例如,定性研究者倾向于回避结构化的问卷和测试;另一方面,定量研究者则常常不愿利用非结构化访谈以及对观察和文件的内容分析所提供的丰富资料。尽管存在着这些隔膜,定性研究路径和定量研究路径在测量的基本原则和目标上仍然是非常相似的。主要的区别或许是:定量测量的基础通常是将事件/属性归类到业已构建的范畴之中,而定性测量的基础则更多的是将之归类到新形成的范畴或解释之内。但是,甚至这一区别也并不总是一成不变

的，因为许多定量研究者可能会在已经预知的范畴（或应对）之外寻找新的范畴（或应对），而许多定性研究者会在对主题进行分类时使用一些预先构建的标准。

测量的量表

根据变量的性质，研究者可以根据一系列“量表”对变量进行测量。最简单的量表形式，即所谓的“定名量表”，是一种将个体和/或事件区分到互相排斥的组群中的量表。这种量表的典型包括根据个体的性别、种族、婚姻状况来区分个体。定名量表是一种性质量表（区分的是性质，而不是量级）。因此，如果对不同的组群赋予不同的数值（如 1 = 男性，2 = 女性），那么，这些数值便仅仅是用作标签，而不具有好与坏的含义。

在定性研究中，情境、事件、个体通常以复杂的叙述形式得到描述表达，这些描述形式是互不相容的，或者说是研究者努力使其互不相容的（如 Lincoln and Guba，1985）。尽管定性研究者并没有明确地指称这些描述和记录为测量，但它们的确是测量的形式。这种形式的测量量表可以被称为“定性—描述量表”。由此，不同于那种将事件和个体分组到定性范畴（如定名量表）之中的做法，这些事件和个体也可以根据多元的性质维度进行分类，并用叙述的形式进行描绘。

当定名量表中的不同层次可以根据其量级或某些定性属性（如吸引力）来加以确定时，测量的量表就成为了“定序量表”。定序量表可以将事件、情境、个体区分成不同等级的类型——对婚姻满意度、主观幸福感以及数学能力的测量都是这种量表的典型。定序量表的一个主要问题是，排序中的同一差距并不意味着实际属性上的同等差距（例如，较之于 70 - 63 = 7、77 - 70 = 7 可能会有很大的不同）。

“等距量表”则解决了这一问题，其办法是将数值赋予不同的层次/量值，进而使相同的赋值之差代表着“真实”属性之间的相同差。例如，85 度与 80 度之间的实际温差相较于 80 度与 75 度之间的实际温差是相同的。等距量表有一个零点值，但那不是“真零点”。例如华氏 0 度并不意味着没有热量。也就是说，如果今天是 80 度，三个月之前是 40 度，即便 80/40 = 2，我们也不可以下

结论说今天比三个月之前热两倍。

为了使这种比率有意义,就需要有一个“真零点”,使之成为可能的测量量表是“定比量表”。这种量表的典型包括对高度的测量和对“兄弟数目”的测量:一个有着 4 个兄弟的人是那些有 2 个兄弟的人的两倍。同样,在给定时间内跑 4 里路的人的速度是在相同时间内跑 2 里路的人的速度的两倍。但是,遗憾的是,很少有教育学、社会学或行为的测量可以应用这种测量量表。

测量量表对于资料分析具有重要的意义,对立基于测量的推论类型也至关紧要。大多数常用的统计程序采用的是等距量表(或接近等距量表)。对于诸如智力、萧条和生活质量(研究对象的幸福感)等变量的测量,都是应用定序量表的范例,因为我们往往认为它们的量值“似乎”是等距的。

测量中的问题:效度和信度

不考虑测量的性质、类型或量表问题,研究者还必须回答与所收集的资料相关的两个基本问题。第一个基本问题是:“我真的是在测量和记录我所意欲测量和记录的东西,而不是其他事物吗?”这是一个测量效度的问题。例如,我们可以假设,研究者所研究的问题是关于校长的“领导风格”的,那么,如果研究者正在设计一份针对教师发放的调查问卷,询问他们对其校长之“领导风格”的看法。效度问题就会询问:“在何种程度上,由此所获得的关于每位校长的分值真实地反映了他的‘领导风格’,而不是反映了其他概念(如社会性、外向性等)?”当研究者在观察某一学校并试图记录下反映校长“领导风格”的指标时,这一问题同样也是重要的。显然,在这个案例中,“概念”(领导风格)的理论含义必须在得出任何结论之前以清晰的形式给出界定。

第二个基本问题是:“假设我所测量和记录的东西是我所意欲测量和记录的,那么,我的测量和记录是否准确无误?”这是一个测量信度的问题。如果测量工具是可信的,那么,无论时间多长(检测—重复检测的信度),无论次数多少(内部一致的信度),也无论评定者/观察者有多么不同(观察者之间或评定者之间的信度),也都应该能一致地提供相同的结论。

关于研究成果的许多不同争论是由于研究者不能以有效、明

确的方式回答上面的两个问题而产生的，在测量效度问题上，尤其如此。因为教育学、社会学、行为研究中的大多数属性都不能直接观察（它们是构建的概念，或隐含的变量。见 DeVellis，1991：12），所以也就无法直接观察出变量（如创造性）与所获得的测量结果之间的对应程度。当不能直接观察他人的创造性时，研究者怎么知道他们是在测量“创造性”？要对测量效度作出判断，就必须以一种可观察的（通常称之为“操作化界定”）方式对概念进行界定。不同研究者的研究成果之所以并不一致，原因之一也许就是他们对同一概念的界定并不一致。例如，一些研究者将“自我意识”（self-concept）界定为一般的形式（如以积极或消极的态度对待自我），或所谓的“自尊”（Rosenberg，1979）。另一些研究者对“自我意识”的界定则是围绕着具体的领域来进行的，如对自己的身材吸引力或认知能力的自我信念（具体的案例可见 Marsh，1994）。比较研究来自不同种族和民族背景的儿童时所出现的结论的不一致，也许应当归咎于概念界定上的不一致（见 Tashakkori，1993）。

确定测量的效度

如果某个概念是无法直接观察的，那么确定观察或测量结果（包括定性研究的“人类工具”）的效度，就其最佳条件而言，也是十分困难的。采用“形似”（looks like）工具去测量政治保守主义（或采用定性/开放式访谈确定某个个体的这一属性），可能会对从同意的需求到宗教的虔诚等一系列属性进行测量。因此，那些“形似”（“表面效度”）工具（或访谈）并不真的就是其所真正测量之物（“概念效度”）的真实指标。表面效度并不能反映工具效度。事实上，研究工具越不明示其所希望测量的对象（“招摇”越少），参与者根据对研究者的目标/问题的觉察而进行反应的可能性就越小（“参与者反应”越小）。

由于回应者的主观评判（表面效度）并未真实体现工具效度，研究者就必须使用其他策略确定其测量效度。在此有两条一般性策略可以采用。一条是对“专家”进行测问，从而帮助研究者判断出一个特定的测量工具在何种程度上对应予测量的对象进行了测量。但由于专家们在许多时候无法达成一致，因而这种效度

类型(“判断式效度”)只适用于那些能够得到明确界定和客观评估的属性。另一个确定测量效度的方法则是收集测量结果的经验数据(“经验式效度”)。

判断式效度(judgemental validation)可用于确定那些对有明确界定的特定属性(如学术能力、学习程度)进行测量的测量工具的效度。通常,这些属性关涉具体的“内容”领域或一组教学目标,评估一次数学考试在何种程度上能够测出学生对课程目标的掌握,就是这种效度(“内容效度”)的典范。显然,如果一个属性的内容没有得到很好的界定,(例如,对于堕胎或某一政治候选人的态度究竟是什么内容?)那么,这种类型的效度就是不合适的。尽管有些作者错误地提及了态度量表的内容效度或其他类似的测量,但这种类型的效度最适宜用来对学术能力进行测量。

经验式效度(empirical validation)是评估测量效度最常用的方法。这种类型的效度通常同时需要两种信息:①体现着某一概念的测量结果与理论上预期有关的其他概念的测量之间所具有的关联(相一致)程度(“聚合效度(convergent validity)”。见 DeVellis,1991:50);②测量结果与预计非相关的其他概念的测量间所具有的不相关(不相一致,不相联)程度(“区别效度(discriminant validity)”或“离散效度(divergent validity)”。见 DeVellis,1991:50)。聚合效度还包括特定的测量结果与使用其他方式测量同一概念所得结果之间的一致性(用定性研究者的术语来说,即多元测量方法的三角测量)。如果说,测量的结果是有效的,那么,这一结果就应当与对其他相关概念的测量、对同一概念的其他测量相一致,并且应当与对其他非相关概念的测量不相关。

所谓“已知组群效度(known-group validity)”(详见 DeVellis,1991:47),就是这种同时性聚合与区别的范例。研究者对那些理论上预计相似或相异的组群进行相互对比,从而确定测量的效度。例如,为了保证对“创造性”的测量效度,艺术家们(按照定义,预期他们会具有很高的创造性)应该比其他组群(如会计,预计他们具有较低的创造性)具有更高的平均分值。

有时,在指标分析(item analysis)时,如果没有合适的标准能够反映某一概念,那么就可以用总的测试分值作为评估测试中每个细项之效度的标准;如果在一组参与者中,一个细项的分值与

总测试分值相一致（如高度的"项目—总体相关性"），那么，这个细项就可以看成是对概念的一种有效测量。单独使用这种细项效度是危险的，应该结合使用至少一种其他的方法。表4.3概述了定量研究中不同的效度确认方法。需要注意的是，尽管有关确定测量效度之常用技术的讨论，往往与定量研究路径相关联，但研究者也可以将其基本原则应用于定性观察。定性观察虽然无法计算相关系数（correlation coefficient），但这并不妨碍我们通过其他方法（如多元资料来源的三角测量）来评估测量的效度。

表4.3 确定测量效度的方法（特别是在定量研究中的）

效度类型	特征和过程
内容效度（content validity）	对一项测验中的各个细项对其所针对的教学目标或内容完成测量的程度，由一组裁判（"专家"）作出评估（但这一方法并不适用于没有具体"内容"的概念）。
同步效度（concurrent validity）	对一群个体进行测量所应有的工具，与对同一概念已具效度的（完全确立的）测量进行比对，当这一新工具所获得的分值与已确立的测量高度关联时，这一新的工具即可被视为有效的。
预测性效度（predictive validity）	当测量可以"预测"出某一概念的具体结果时，这一测量即被视为是对这一概念的有效测量。
建构性效度（construct validity）	这与一项测试对概念完成测量的程度有关。由于这一概念不能得到直接的观察，因而对这一概念的有效测量应当同时达到以下两个条件：①这一测量结果应当与对同一概念的其他测量，或与对那些理论上相关的概念的测量，具有很强的关联（聚合效度）；②这一测量的结果应当与另一测量（即对所关注之概念预计无关的概念的测量）之间不存在关联，或者这一测量的结果应当与已知具有不同程度的概念群体中存在着差异（区别效度）。

确定测量的信度

信度是指测量结果能够准确地反映概念之真实“量级”或“性质”的程度。由于我们无法直接观察这些概念,因而确定测量结果的准确程度并非易事。但是,仍有一些间接且通常有效的方法可以确定对属性进行的测量的准确度(或误差大小)。大多数评估测量信度的方法都隐含着两个假设。一个假设是,如果说一次测量是准确的,那么这一测量在时间上就是可以重复的,或者是可以用相同的测量方法来获得的(如平行测试、第二观察者等)。这实际上是一种在同一组群或同一情境下同时使用两次或多次相同方法的三角测量方法,或一种在更多情境下使用相同方法的三角测量方法。

另一个假设是,如果测量存在着误差(随机误差),那么,在反映属性之“真实”大小或性质方面所产生的这些误差会在重复测量中彼此抵消。例如,如果一个观察者倾向于对某一属性给予超出其应得的更为积极的评估,而另一观察者则可能会在对同一个人的同一属性进行评估时多少倾向于消极,那么随着观察数量(或测试中的项目数量)的增长,测量“真实”属性的误差也就趋近于零。表 4.4 总结了对测量/观察的信度进行评估的不同方法。需要再次强调的是,尽管这些确定测量/观察之信度的技术通常是在与定量方法相关的意义上所展开讨论的,但其基本的原则也可以用于定性观察。我们在表中已经指出了这类相似之处的部分内容。

表 4.4 确定测量信度的方法(特别是在定量研究中的)

信度类型	特征和过程
重复检测法	如果重复进行检测的结果能够以一致的方式区分某个组群中的成员,那么这一检测就是可信的:通常是通过计算相同个体组群的两次检测之间的相关系数来进行评估。
分成两半法	两半检测所获得的结果之间保持一致的程度:通常是通过计算两半检测的得分之间的相关程度,以及校正检测的长度(Spearman-Brown 校正)来进行评估。

续表

信度类型	特征和过程
平行形式法	两项平行检测所获得的分值之间保持相互一致的程度:通常是通过计算在一个个体组群之中两次检测的分值之间的相关程度来进行评估。
内部一致法	测试中的不同细项以一致的方式测量属性的程度:通过计算测试中的不同细项之间的平均相关程度来进行评估。Cronbach 的系数 α 与 Kuder-Richardson 的公式 20 和 21 就是这种类型信度的范例。
判定人或观察者交互法	两名或多名评定人的评分之间,或两名或多名观察者的观察之间保持相互一致的程度:可以通过计算两名评定人对一群个体的某项属性所给出的一组评分之间的相关程度来加以确定;对于定性观察而言,可以通过评估两名观察者在观察同一环境下的同一现象时的同意程度来加以确定。
判定人或观察者自省法	一名评定人所作的两个或更多的评分之间,或者一名观察者所作的两次观察之间保持一致的程度:通过计算两组分数之间的相关程度,或者通过主观比较不同的定性观察来确定一致程度。

注:关于这些以及其他一些确定测量信度的程序,详见 DeVellis(1991)。

推论/结论的质量:定量研究的视角

通常情况下,在研究过程的最后,研究者会根据对资料/信息的收集和分析推导出关于各个变量或事件之间关系的结论或推论。在提出这样的推论时,面临的主要问题就是对这些关系会提出其他替代性解释的可能程度。在上文中,我们已经从推论或发现的内部效度方面对这一问题进行了讨论。而推论正确性(即内部效度)的置信度,则取决于对这一结论是否还存在着其他替代性解释。我们在前面已经讨论了对内部效度构成威胁的一些因素。

例如,在本章的开头,我们已经讨论了选择性(selectivity)对研究发现的置信度的影响。研究者所观察到的关系也许源自于这样的事实,即一组个体或事件在与其他个体或事件就特定属性(如一个自变量或预测变量)进行比较时,这两组个体或事件在该特定属性之外的其他属性上也是不同的。例如,在收集了详细的资料后,研究者可能会得出这样的结论:受过学前教育的孩子在小学中会有更高的智商和学习成绩(见 Tashakkori, Haghighat and Yousefi, 1990)。这一研究结论是在对这些孩子与那些小学一年级前未接受过学前教育或未上幼儿园的孩子进行比较后得到的。但研究者也许会猜测,这些推断出的差异也可能源自于拥有这种经历的孩子出于不同类型的家庭背景(在社会经济地位、社会心理特点等方面的意义上)。由于对这一结果存在着其他可能的解释(或替代性解释、竞争性解释),因此,所谓学前教育经历造成学童差异的结论,其信度或内部效度就很低(或为零)。当然,的确有一些方法可以用来控制、检测、排除这类替代性的解释。下文即是对其中一些办法的讨论。

表4.5说明了一些可能降低结论或推论之质量(损害内部效度)的因素(也可参见:Campbell and Stanley, 1966)。除了选择性问题,我们也已经讨论了对“统计结论效度”造成威胁的一些因素来源,如小规模样本,以及因测量可靠性低所导致的随机变异(误差)(其他威胁因素可见:Cook and Campbell, 1979:39-50)。对结论质量的大多数威胁,来源于对可能影响到研究者所关心之变量的“外部变量”的控制不足。我们在第3章中对这一问题进行了讨论(即“最大最小控制”原则中的“控制”原则)。对于实验型研究而言,保证所有变量和事件在实验群体和比较群组上的相同性,便可消除大多数的这类威胁;对于非实验型研究而言,小心仔细地检查相对于研究者的发现是否还有其他可能的解释,或研究者的变量之间是否还存在着其他关系,并且随后对这些解释的可行性进行评估和讨论,就会提高研究者之推论的内部效度(信度)。这样的评估既要在理论上展开(即评估在理论上和此前研究的基础上何者为可行),也要通过对资料重新分析以评价其他可能的关系。

表 4.5　一些损害研究发现之推论质量(内部效度)的威胁

威　胁	特　征
历史	研究过程中的事件可能会影响一个组群,但不会影响另一个组群,这导致组群之间的差异并非只受自变量的影响。在非实验性研究或定性研究中,历史是指在研究者正在研究的事件之前和之上发生的事件(对一群个体)。
选择	在研究开始之前,一个组群的某些属性即已区别于另一个组群,由此,在研究处理之后所出现的差异(或在非实验性研究和定性研究中的特定事件)并不能完全归因于自变量/标准变量。
统计回归	当研究者依据某一极端属性(如高绩效或低绩效)而选取研究对象之后,预先测试和事后测试之间所出现的任何差异都有可能源自于走向次级极端值的(随机)趋势(如果存在随机变异,处于极低分数的学生的分数还能往下滑动吗?)。同样的威胁也会加之于对既有的极端个案/群体进行的非实验性研究或定性研究。
成熟化(maturation)	预先测试和事后测试之间的差异可能源自于参与者身心的成熟化,而非自变量的不同;而且,两个组群之间的不同也可能源自于一个组群的变化(即成熟化)节奏与另一组群的节奏不同(选择—成熟化之间的互动)。
预先测试	预先测试和事后测试之间的差异可能是由于个体对测试的熟悉(重复的效果)所产生的,而不是自变量的不同所导致的。
工具化	预先测试和事后测试之间的差异(或无差异)可能是由于测量(测试、观察)的随机变异(缺少信度)而导致的,而不是由于自变量/标准变量的不同而引起的。
主观磨损	预先测试和事后测试之间(或两个组群的分值之间)的差异可能是由于一个组群中漏掉了一些个体所导致的。
执行	研究所获知的变量之间的关系可能归因于:①实验者/研究者/观察者的预期;②正待研究的参与者的反应行为。

注:详见 Ary 等人(1996)。

对于那些损害推论之质量的可能威胁因素,可用于降低或消解的策略列举如下:

1. **随机分出组群。**这是创建在所有方面都较为相似的、两个或更多组群的最有效的方法。由于回应者被随机地分配到各个组群之中,因而各组群之间预计不会在外部变量上存在系统的差异。这只适用于实验型研究。

2. **配对。**如果事先已经辨认出特定的威胁变量,那么,研究的参与者便可以根据这些属性进行配对。成对的每个个体被分别随机地分配到一个实验组群中,由此产生在被配对的变量上高度相似的两个组群。这一方法可用于实验型研究。在大多数这类应用的案例中,样本是随机抽取出来的,形成匹配好的两个主体(或三个主体),然后将之随机分到不同的组群之中。在因果比较研究和定性研究中,选择对个体、观察单位进行配对是可能的。例如,在每个定性研究的案例中,会发现有一个在特定属性(如性别、教育水平)上相似的比较案例。显然,在这一应用中,采用的是在选择上进行配对,而不是在分配上进行配对。

随机化的配对降低了组群间在一个或多个外部变量上出现差异的概率。但配对也可能会导致主体的丢失(并削弱外部效度),因为某些主体会缺乏配对者,特别是在对一个以上的变量进行配对(如教育和收入)时,尤其如此。

3. **同质分群。**如果预计有一个外部变量(如性别、种族、教育)会扭曲研究中的主要变量间的关系,那么,就可以根据这一属性划分样本,而此后的资料分析也在这些组群中各自分开进行(如在男性组与女性组分开进行分析)。对某些定量研究来说,更先进的分析方法是使用一种"因子设计"(factorial design)。在因子设计中,外部变量被看成与研究中的主要自变量一样的因子(一个自变量),这样就可以为检验外部变量和自变量对因变量的交叉影响提供一个机会。

4. **个体自相比较。**当个体的属性可能威胁到推论的内部效度时,可以将每个个体与其处理前后的自身进行比较。在相关的研究文献中,这类应用的典范包括事前检测—事后检测设计、重复测量设计、主体之内研究设计、时间系列研究设计和单个主体

研究设计。在引入某个干预变量或事件之前和之后的定性研究中,也可以应用这一方法。

5. **统计控制**。在控制外部变量之“影响”的其他可能性不复存在时,最后的策略或许是在统计上将与这些外部变量相关的变异剔除掉。“协方差分析”(analysis of covariance,ANCOVA)就是其中一种方法,它可以用来在对群体均值进行比较的情况下,控制外部变量(协变)对因变量的影响。分割关联(partial correlation)是另一种方法,可以用来计算当第三变量(外部变量)在统计上被控制住(分割出去)时,两个变量之间内在的关系。还有一些更复杂的统计程序可以用来达到这个目的,如多元回归分析、路径分析(path analysis)等。在统计控制之后,估测变量间关系的强度或效应影响的重要性等都是这种策略应用的重要方面。(关于量级对统计显著性的一般性讨论,参见 Maruyama and Deno,1992:112-113)

6. **双盲程序**。这种策略用来解决在研究实施过程中所出现的威胁因素(即“研究者的预期”与“参与者的反应性”,见表4.5)。为了避免参与者的反应性行为,必须使参与者不能觉察到研究课题的确切性质;为了降低研究者的预期性,必须使实验者、访谈者或观察者忘记研究课题的主要预想或假设。显然,如果研究者同时也是程序处理的实施者、观察者、资料收集者,或定性资料/观察的内容分析者,那么,要做到这点是不可能的。例如,在比较研究教授阅读的两种方法的相对效率时,如果研究者同时就是教师,那么他的预期就可能会影响研究结果;当教师主动请求根据自己对方法的喜好而使用特定的教授方法时,也会影响到研究结果。于是,在这些学生(在特定方法下接受教育)中所观察到的差异,就可能更多地出于教师的人选,而不是教授的方法。

在本章的下一节,我们还将考察和评论一些对威胁定性研究之推论质量的因素给予应对和处置的平行策略。上面所提及的大多数策略也可以用于定性研究,但本节中关于定性研究所提及的许多策略都可以直接地、或部分修正地用于其他类型的研究之中。作为混合方法的主张者,研究者可以应用两组策略之一或两组策略的融合而提升推论或结论的质量,从而受益无穷。

资料的质量与推论:定性研究的视角

我们此前已经论及,定性研究者尚未认真地对测量问题(测量效度和信度)与推论质量(设计效度和内部效度)作出区分。因此,我们在本节中将同时对这两个问题进行讨论,但我们仍将集中论说那些与确定推论质量更具相关性的策略。

"可信性"(trustworthiness)是 Lincoln 和 Guba(1985:290)提出的一个已经得到公认的定性研究的概念,用来指代、喻示许多定量研究的设计问题和测量质量的问题:

> 与可信性相关的基本问题很简单:研究者如何能够说服其听众(包括他本人)相信他的研究发现值得关注、值得考虑呢?安排什么讨论、使用什么标准、提出什么问题,才能在此议题上具有说服力呢?

Lincoln 和 Guba 随之介绍了四个标准,即信用性(credibility)、可转移性(transferability)、可靠性(dependability)和确定性(confirmability),它们集合起来可以确定一项研究的可信性。下文所列举的前 7 种方式能够使定性研究者建立关乎其研究发现的"信用性"。其中,已经有几位理论家对信用性(如 Eisner,1991;Janesick,1994;Lincoln and Guba,1985;Patton,1990)进行了详细的讨论,因为信用性是定性研究的结论和推论建立其可信性的最重要的成分。下文即列举出了确定定性研究之可信性的不同方法:

1. **延长在场时间**。花费足够多的时间去建立信任、了解"文化",并检测出源自信息提供者或研究者自身之偏见的错误信息,这对研究者来说是非常重要的。本书第 1 章已经指出,Margaret Mead 之所以被批评者指责她所报告的萨摩亚人信息提供者给出的信息是有误的,在相当程度是因为她在聚焦到特定的研究领域之前,没有花费足够的时间去了解萨摩亚的总体文化。延长在场时间的目的在于,使研究者能够意识到在任何既定的社会场景中都存在着多样性的情境因素以及信息提供者的多重视角,进而使研究者能够确定研究课题的"范围"。应当注意的是,这一方法更多地关涉信息的质量,而非立基于这类信息之上的研究推论或结

论的质量。

2. **持续观察**。持续观察的目的是帮助研究者充分辨识与其所研究的问题最为相关的社会场景的特点或方面，进而为研究者提供研究的“深度”。这一研究活动同样关涉信息的质量远多于对推论或结论之质量的关注。

3. **使用三角测量技术**。本书第 1 章和第 3 章已经描述了这类方法，包括研究资料的三角测量（如访谈和观察），研究方法的三角测量（如定量方法和定性方法），以及研究者本身的三角测量（Denzin，1978；Patton，1990）。

必须提请注意的一个忠告是，许多定性研究的理论家和研究者并不使用三角测量，因为他们认为并不存在可以应用三角测量的单一事实。Denzin 和 Lincoln（1994）在介绍 Richardson（1994）的定性研究取向时，将之称为晶状体而非三角形：

> 传统的后实证研究依赖使用三角测量，包括使用多元方法，以此为确证效度的方法。这种模型所隐含的假设是，这里存在着一个可以得到三角测量确定的固定参考点。后现代主义者和混合流派的文本并不进行三角测量，其核心景象是一个晶状体……晶状体是一种可以通过反射和折射，创建出映照现实的不断流变的景象和图画的棱镜。结晶化解构了有关效度的传统观念，因为现在已不可能有单一的、可被三角测量的事实了。（Denzin and Lincoln，1994：482）

当然，这并非本书所认可的研究取向。在本书中，我们的看法是，通过促进混合方法的使用，进而可以得到具有所谓内部效度或可信度的结论和推论。

4. **同行评议**。这是一个过程，是“以一种类似于分析进程的方式向中立的同行们展现自己的研究，旨在探讨研究进程的各个方面，以免这些问题仍只是含混地隐藏在研究者的内心之中”（Lincoln and Guba，1985：308）。这一过程对增强研究的可信性是有益的，因为它使研究者充分地展示在那些试图发现偏见的同行们面前，并向他们澄清阐释性问题。无疑，这对于提高研究者在资料分析之后所作出的推论/结论之质量是很重要的。也正是因

为这样,这一过程更接近于此前所讨论过的"内部效度"的理念。下文(第(5)、(6)条)所述的两个行动也同样如此。

5. **反面案例分析**。这是定性研究中相当流行的方法(如Glaser and Strauss, 1967; Kidder, 1981; Lincoln and Guba, 1985; Patton, 1990; Yin, 1994),对此,我们可以描述为"在辨识出模式和趋势之处,我们对这些模式和趋势的理解将因省思那些有悖于这些模式的情况和案例而大为增进"(Patton, 1990: 328)。反面案例分析的目的是,修正所建模式直至相互矛盾的案例也能符合所建模式,从而最终消除反面案例。

6. **充实参考资料**。这一方法是在档案中储存某些原始的定性材料,以便以后的再回顾和再分析(如见 Eisner, 1975, 1991; Lincoln and Guba, 1985)。最初建议使用的是录音磁带和摄影记录,现在已经扩大到所有形式的定性材料。这里的再分析可以由原来的调查人进行,也可由其他分析者进行。新的阐释可以通过与原有阐释的对比衡量强化其信用性。

充实参考资料与统计分析之间可以作出直接的类比。例如,使用回归分析的研究者有时会将其样本分成两半,在第一半样本上进行分析以获得解释,随后在第二半样本上进行确证分析。同理,定性分析与定量分析在这里依据的逻辑都是一样的:只是一个使用某种形式的文本(如文字、图画),而另一个则使用数字。

7. **成员自查**。这往往是在研究过程期间或在进行总结时展开,并构成信用性审核最为重要的环节。一些著者已经描述了这一过程(如 Lincoln and Guba, 1985; Spradley, 1979),其方法是对参与研究社会场景的成员进行询问,以检查研究者所使用的分析范畴、结论和解释。Spradley(1979)主张使用结构化问题,让受访者对研究者已经建立的分析理论进行确认,如果受访者或听众同意研究者所作的阐释,那么这就为结论的信用性提供了确认证据。

8. **深描**(thick description)。这一方法能够为定性研究的阐释和结论的可转移性提供支持证据。如前所述,定性研究者喜欢将推论从一个特定的发送背景转移到另一个接收背景之中,而这就需要对所有信息进行详细的描述(即深描)(Geertz, 1973; Lincoln and Guba, 1985)。这一概念接近于一般研究、特别是定量研究中

的推论或结论的“外部”效度的概念。

9. **可靠性审查**。Lincoln 和 Guba(1985)对可靠性审查进行了总结,认为这一审查关乎研究的全过程,包括研究决策和方法论转变的适切性问题。这一活动可以归入更具一般性概念——“转达的忠实度”之中。对此,前文已有论述(Krathwohl,1993)。

10. **确定性审查**。Lincoln 和 Guba(1985)曾给出结论,确定性审查所关注的是研究的成果(即要证明研究的发现和阐释是否能得到资料的支持,是否能做到内在自洽)。我们将这一过程称为“推论一致性审查”,即确定推论的内部效度。

11. **自省记录**。这一方法是要为可信度所有的 4 条标准(信用性、可转移性、可靠性和确定性)提供信息。Lincoln 和 Guba(1985:327)将自省记录描述为:

> 日记的一种,研究者逐天或依需而作,将与自我……和方法有关的大量信息记录下来。在自我方面,自省记录可看成是提供关于人类工具方面的同类资料,在常规研究中这方面的信息总是关于纸笔或黄铜工具;在方法方面,自省记录提供的信息包括如何作出方法论选择,以及选择的理由——这些信息对审查者同样是非常重要的。”(原文为斜体字)

作为对本章的最后总结,我们认为,尽管确定定性结论的“可信性”与确定定量推论的“内部效度”的“活动”类型未必相同,但定性和定量这两种方法在认定“推论质量”是观察/资料之结果的看法上却是高度相似的。由此,一个包含了两种方法之要素的融合性框架就是完全可行和可能的。我们在本章中已经尝试指出这些共同点与相似处。我们希望,我们将来能够创建出一种“评估研究发现和推论的超级框架(a superordinate framework)”。

5

资料收集策略和研究程式

资料收集的传统路径和替代路径

资料收集工作的传统路径有一个显著的特征，即研究者对于定性研究路径或者定量研究路径的“忠诚”不二。尽管仍有忠诚如斯，但越来越多的研究者们开始在一项研究中同时运用定性研究路径和定量研究路径。不同学科背景的研究者已经开始倡导运用这种“多元”的方法来收集资料。譬如，Sechrest 和 Sidani（1995）提出，在研究的 4 个阶段（即资料收集、数据分析、阐释和运用）中，任何一个阶段都可以同时运用两种方法：一种是“公式的”（formulaic），另一种则是“临床的”（clinical）。他们所说的“公式的”资料收集包括问卷、结构化访谈和观察程序等传统的焦点/结构化测量程序，而“临床的”资料收集则主要涉及临床式访谈、非结构化观察以及对文献的整体性评估等。

所有这些资料收集方法，往往能够给我们提供关于研究对象或情境的全新且未经探查的信息。但单独使用其中任何一类资料收集方法，都可能会造成资料的不充分或偏颇。通过这两类路径的融合使用（Sechrest 和 Sidani 称之为多样方法或互补型方法）而获得的资料，要比单独使用任何一类路径所获得的资料都丰富。其中的一类方法证实了或补充（加强）了经由另一类方法所收集的资料。

互补型方法，也就是我们所说的混合方法，可用于同一次测

量之中。例如,在一次“漏斗式访谈”(funnel interview)中,研究者可以先提出一些宽泛的、开放的问题,以便从受访者那里引导出不落窠臼的信息,然后再提出事先编制了回答模式的一些封闭性的、结构性的问题。另一种更为常用的方法是,研究者可以在问题之后列出事先编制的回答,后面再为“其他”回答选项提供一个空白处,而答卷者可以坦率地在这一空白处写上自己另外的评论或提供另外的信息。对于这些“混合的”或“互补的”资料分析程序的各种可能性,我们将在下一章进行讨论。

尽管大多数关于研究方法的教科书都将定性资料收集程序和定量资料收集程序区分开来,但我们并不这样处理,原因至少有三:①大多数资料收集程序(如访谈)都可以在某一研究或情境下同时使用这两种资料收集路径;②几乎所有的资料收集形式都能从多种方法的混合使用和互补使用中受益;③即便是传统的资料收集模式,亦可归入以高度结构化/定量为一端、以高度非正式性/定性为另一端的连续谱之中。我们不打算在两种传统路径的二择一的框架内来讨论不同的资料收集程序,相反,我们将对每种主要的资料收集程序或模式进行单独的讨论,并指出在这一模式下互补地或混合地收集资料的可能性。譬如,在提到作为资料收集形式的个人访谈时,我们将同时讨论个人访谈中的开放性(定性的)方面和封闭性(定量的)方面。

有策略计划的资料收集

资料收集的传统定量路径,是指使用相对详尽说明的和事先计划的工具进行资料收集工作。另一方面,大多数的传统定性研究,在操作中并没有使用这种事先计划的方法来收集测量/观察资料。但是,大多数的定性研究者都至少有一套关于使用何种类型资料收集方法的策略(即观察、访谈等),甚至有一套关于聚焦观察何种行为和(或)描述点的策略。用 Miles 和 Huberman (1994:35)的话来说:

> 至少从一开始,知道了研究者自己想要弄清的东西之后,也就必然走向了如何获得有关信息的问题……如果我想弄清嫌疑犯是如何被逮捕和被指控的信息,我就会决定去对那些与此有关的人们(警察、嫌疑犯、律师)进行访谈,收集与

框5.1 Miles和Huberman对支持和反对定性研究使用计划性工具准备之意见的总结

主张事先毫无工具准备的观点

1. 预先设计和结构化的工具会使研究者对研究现场闭目塞听。如果研究田野实地中最重要的现象或运作的深层结构没有在预先设计的工具中有所反映，研究者就会忽略或错认这些现象和结构。

2. 事先的工具准备往往是脱离环境的，它追求的是普适性、统一性和可比性。但是，定性研究是通过对环境的观察而获得鲜活的生命和呼吸，是由特殊性产生普遍性，而非相反。

3. 许多定性研究只涉及单一个案，其中涉及的人员也很少。那么，难道还需要问卷、观察计划、测验等一类通常用于生成成本低廉、可供比较和参数性分布的大规模样本的工具吗？

4. 田野工作最重要的是做笔记、录音（交谈和会晤）和收集物品（文献、产品、人工物品），因而事先的工具准备（instrumentation）是一个不恰当的名称。一些导向性的问题、观察的标题，以及一个大致准备好的材料分析，足以满足研究者在着手研究时的需求，甚至足以满足研究者在整个研究过程中的需要。

主张进行事先工具准备的观点

1. 倘若研究者已经明了自己所寻求的研究目标，那么，对如何收集信息不进行事先的计划则是不可理喻的。

2. 如果对访谈计划或观察计划不进行聚集，那么，就会收集大量不必要的信息，而过泛的资料将会削弱分析的效率和力度。

3. 使用先前的研究中已经使用过的同类工具是我们进行跨研究课题对话和交流的唯一途径，否则，研究就是不可比较的，除非是非常宽泛意义上的比较。因此，我们需要用共同的工具来构建理论、改进解释或预测，并为实践提供建议。

4. 有偏见的、不了解情况的研究者只会询问片面的问题，作出选择性的记录，进行不可靠的观察并扭曲信息，研究资料因而将是无效的、不可靠的。善于使用有效的工具能够最佳地保障研究发现的可靠性和价值意义。

来源：Miles和Huberman（1994:34）。

> 逮捕有关的文件(如规则条例、记录等)……即使采用的是开放性的访谈或观察,也必须作出一些技术性选择:要作记录吗?何种形式的记录?要对整个互动过程进行磁带录音吗?事后再听?是否转录?如何作出书面记录?

从本质上说,两种传统方法中的资料收集工作都可视为从非计划性的一端到高度计划性的另一端的连续谱。连续谱的一端是偶然性的资料收集活动,事先对观察什么、收集什么资料、如何去做都没有任何想法。但很难想象,在研究过程中会有这样一种完全无计划的观察或资料收集策略。大多数"无计划的"资料收集工作也至少在某些方面是有计划的。在连续谱的另一端,则是高度结构化的测量,例如本章后面所讨论的个性问卷和态度量表(attitude scales)。Miles 和 Huberman(1994)对有计划的测量/工具的支持和否定意见进行了回顾总结。框 5.1 即展列了这些论点。

我们认为,资料收集工作应该尽可能地根据研究问题而加以设计。对于某些研究问题而言,只能进行一些有限的准备;但对于其他大多数研究问题,通常都有充分的先在知识,可以使研究者能够形成更有计划性的策略。且不论事先准备的可能程度或可欲程度,混合模型的资料收集方法可以为研究者提供同时使用两种策略的便利。

资料收集对参与者行为、反应的影响

资料收集过程中的一个主要困境是有关个体对其被研究的意识(即参与者反应)。人们基于这种意识而作出的各种反应,都会影响到资料的准确性和研究发现的内部效度或可信度(详见 Rosnow and Rosenthal,1996)。与其他社会交往情形相类似,参与研究的人们也扮演着特定的角色。一个个体也可能会因其意识到作为研究的参与者而扮演多种多样的"角色"。可能的角色包括:①"好心的"或"有用的"参与者;②"担忧的"参与者;③"忠诚的"或"老实的"参与者;④"怀疑的"参与者;⑤"敌对的"参与者。框 5.2 对这些不同角色逐一作了简单的介绍。尽管大多数关于研究方法的教科书会在与实验方法有关的视角下对这些参与者角色进行讨论,但这些角度实际上也适用于所有其他类型的研究。

框 5.2　问卷调查中研究参与者可能扮演的一些角色

“好心的”参与者:这种参与者会努力使其行为合乎其所猜想的研究者之期望,进而对研究者提供“帮助”。这种类型的参与者总是试图帮助研究者确证某种假设和(或)预测。

“担忧的”参与者:这种参与者关注于研究者对他(或她)的感受,为避免产生不好的印象,他(或她)会对自己的反应和行为进行控制。这种类型的参与者往往倾向于以适合社会期待的方式行事。

“怀疑的”参与者:这种类型的参与者不相信公开表达的研究宗旨或研究者的目标,因此,他们会以一种与那些相关指示完全无涉的方式采取行动。

“敌对的”参与者:与怀疑型参与者相似,这种类型的参与者怀疑研究者本人或研究的目的。出于这种怀疑或其他原因,参与者的反应是以与其所认定的研究者之预测相反的方式采取行动。

“忠实的”参与者:这是研究工作的理想参与者。这种类型的参与者试图以一种“现实的”和“真实的”方式作出反应或采取行动,而不论其本人对研究者或研究的预测/期望作何理解。从本质上说,这种参与者忠实于真相,而不是自己对调查的感受。

通常情况下,确定某一参与者在研究中所扮演的准确角色是不可能的。很显然,研究者会希望研究中的参与者能尽可能如实地作出反应,以一种像他们未被观察/研究时那样自然和非控制的方式作出表现。只提供对研究工作具有重要意义的信息,求助于参与者的利他主义感觉,创建出对于专业人员的信任气氛,这些都会减少故意扭曲信息和控制行为的可能性。但是,这些策略并没有排除研究参与者的无意识反应的可能性。在资料收集工作之后(或实验之后)对参与者进行访谈通常是可行的,这可以提供关于这些角色的可能影响的信息。研究者可以对所有的参与者或其中一部分样本进行询问,了解他们对研究者的研究工作和他们参与其中的感觉、想法和态度,此外还应当询问参与者对这一研究的主要目标的“猜测”,以及这些猜测可能影响到其行为和反应的程度。

在这些反应性的问题上,令人欣慰的是,在良好操作的研究中,参与者很少会对研究工作或研究者的目的和期望形成“统一”的观点。统一的反应(偏见)由此遂不存在。还有,大多数社会科

学研究都具有多面性,涉及多个变量之间的交叉网络关系。这种研究中的参与者也就不可能察觉到研究的期望和预测,乃至研究的主要目的。

毋庸置疑,一切都取决于研究者在多大程度上能避免透露自己的主要研究目的①。在行为研究或教育学实验中,确有必要通过提供令人信服但却有误的信息以减少参与者对研究期望和目的的察觉。然而,一旦完成资料收集工作,就应当尽快告知参与者有关研究的"真实"目的(关于这一"欺骗"问题的讨论,可见 Greenberg 和 Folger,1988)。

资料收集策略

一般而言,研究者可以使用多种方法(或模式)从事资料收集工作。尽管向人们询问与其本人或他人相关的问题和态度是应用最为广泛的资料收集方法,但这一方法也有其不利之处。其他一些资料收集方法则试图通过对行为的观察、对人造物的分析,以及对个人创造的产出品的剖析,来收集关于个体之行为和态度的信息。此外,研究者还可以在研究中使用其他人或机构(如学校)所收集的信息,或个人所创建的信息(如日记)。在此,我们将对资料收集程序的一般类型进行讨论:

1. 询问个体以获取相关信息或有关其经历的信息;
2. 观察和记录人们在做些什么,并作出推断;
3. 询问个体与他人的关系;
4. 使用他人收集或记录的资料。

询问个体的相关信息和个人经历:自述报告方法

自述报告(self-report)。自述报告在传统的定性研究和定量研究中是最常见的信息来源。这一方法是对个体进行询问,让个体本人报告其感受、信仰、态度和其他特征。我们所讨论的与"反

---

① 需要注意的是,在某些定性研究中,参与者在整个过程扮演着更为积极的角色,充当起了研究者的合作研究者。

应性”(reactivity)有关的大多数问题,大都适用这一资料收集方法。另外,还应当注意“回答定式”(response sets)(即以一定方式进行回答的惯性)问题。这类回答定式的一些偏好包括“默认的偏好或赞同”(即总是“说是”或表示赞同的倾向)、“极端化的偏好”(倾向于选择反应量表的两端)、“折中的偏好”(倾向于选择反应量表的中间值,或倾向于“平均值”)和“正面的偏好”(选择反应选项中的正面选项)。在回应者中如果某个特定群体存在着这类偏好而其他群体却没有时,这些偏好便尤其具有破坏性。群体之间所观察到的差异可能就是这些不同反应倾向的结果,而不是任何“真实的”差异的体现,而根据这些结果所作出的任何推论或结论,其内部效度(即推论的质量或可信性)将会很低。这样的案例可以参见 Marin 与 Marin(1991,第 6 章)对拉丁裔受访者之偏好倾向的讨论,对这些倾向和其他反应倾向的一般性讨论,可参见 Mangione(1995:33-36)。

访谈(interview)。作为一种资料收集的方法,访谈在定性研究中比在定量研究中更为常用。尽管方法论教科书会将访谈分类为定性访谈和定量访谈,但实际上这里存在着一个连续谱,从非结构化的、开放的访谈到高度结构化的、封闭的访谈:传统的定性访谈位于连续谱的一端,而传统的“定量”访谈则位于连续谱的另一端。定性访谈往往是非直接的,比较笼统(如“请谈谈你的学校”);而传统定量研究中的访谈则更为结构化,通常是封闭性的(如“你认为以下哪一选项描述了你们学校食堂的食物:‘非常好’,‘好’,‘差’,‘非常差’?”)。

在传统的定量研究中,开放性的访谈也应用于问题研究的早期阶段,此时还没有获得相关的研究信息。这在跨文化和多元文化的问题研究中尤为关键,因为研究者对于研究对象总体的心理报告系统并不十分熟悉。例如,在建立态度量表的过程中,研究者会使用开放性访谈来获得与研究对象相关的评价。这些评价随后用初步的“项目群”(item pool)的形式进行组织化处理。然后,研究者再将这些项目用在结构化的访谈或问卷之中,从而对人们的态度和观点进行测量(李克特量表,瑟斯通(Thurstone)量表等)。

在传统的定性研究中,非结构化的访谈在个体情境或群体情

境（如“焦点小组访谈”。见 Stewardt 和 Shamdasani，1990）中都得到了广泛的应用。通常，研究者会对访谈对象的反应进行记录（录音、录像或二者兼用）和登录，并对之展开“内容分析”（我们将在本书第 6 章讨论其间的各个步骤）。

尽管一些作者将开放式访谈和封闭式访谈作截然两分，但许多研究访谈都将这两种访谈形式结合起来使用。其中，最常见的一种结合方式就是“漏斗式访谈（funnel interview）”。在这种访谈中，研究者从十分宽泛的问题开始询问，然后逐渐将问题的范围缩小到少数焦点议题之上。这种类型的访谈无疑就是混合研究路径的应用。

访谈是一种强大有效的资料收集方法，它为研究者（或资料收集者）提供了与被研究个体（或焦点小组中的部分参与者；见 Krueger，1988）一对一的交流，也为研究者提供了可以请求访谈对象就模糊答案进行澄清或对不清楚的问题进行澄清的机会。开放式访谈更可以就访谈议题产生丰富的信息，这些信息也许会引导研究者以一种与最初的预想完全不同的方式对议题进行概念化。

访谈也有较大的缺陷和劣势：耗时多、花费大。如果没有外部资金的支持，很少有学位论文的研究项目能够进行大量的访谈。但通过综合使用电脑和电子通信设备进行访谈，可以在一定程度上解决这一问题，关于这种运用的详情可见 Lavrakas（1993）。访谈的另一个主要缺陷是存在着访谈者影响访谈对象的风险：研究者可能会在不知道的情况下通过姿势、习惯或语言反馈影响到受访者的反应。如果研究者同时又是访谈者时，其更会在访谈过程中对符合自己预期的表述或反应显露出赞同的微妙迹象。在电话访谈中，这种影响相对较小，但电话访谈者又会通过语言影响到受访者的反应。有关减少这种影响的策略，可见 Lavrakas（1993）。

在访谈涉及的议题中，研究者还可以加上前面讨论过的“反应性”问题。此外，研究者对所获信息的记录也可能会受其本身所预期的影响，特别是在研究者必须根据自己在访谈过程中的观察而对受访者的特性（如性格、动机等）作出推论之时，尤其如此。

最后,对开放式访谈所进行的分析要比对封闭式访谈或调查问卷所作的分析更为费时、耗资。关于缓解这些问题的策略,见 Fowler 和 Mangione(1990,尤见第 9 章)。

尽管存在着上述缺陷,访谈仍往往是对读写能力不佳的个体(如儿童)进行资料收集工作的最佳方法。而且,在需要对议题和关系问题有深刻认识的情形下,(例如在新的环境中或对不熟悉的文化群体进行研究时)访谈是最适用的方法。跨文化和多元文化研究就是这种应用的主要典范。

调查问卷。问卷调查是主要采用纸笔进行资料收集的方法,在个人计算机出现后,如今也广泛地使用电脑来陈述问题和记录回答。基于研究问题的复杂性,无论是纸笔式问卷还是电脑式问卷,都要求有一定程度的复杂性和阅读能力,而这可能是被研究的个体所不具备的。在大规模的调研中,调查者有时会向少数读写不佳的访谈对象"朗读"问卷。

调查问卷的一个优势是研究者可以将它们邮寄给相关研究对象。较之于访谈或由研究人员操作的调查问卷(personally administered questionnaires),邮寄问卷调查操作起来要便宜得多。但是,对于未反馈问卷的对象,研究者通常必须随后使出各种办法进行催促,包括两到三次地予以提醒,并且至少再邮寄一次或两次问卷给他(关于邮寄调查的系列步骤,可见 Ary 等编,1996;Mangione,1995)。但即便付出了最大努力,一些人也从不给出反馈。研究者原始样本的这一"损耗",可能会威胁到研究结论的通则化(即外部效度)和推论质量(即内部效度),特别是在不反馈者与总体剩余的其他人之间存在着系统化区别的时候(如某一种族、某一信仰群体),尤其如此。对有关这类损耗和不反馈等问题的讨论,以及对减少和消除这些问题之可能办法的讨论,可见 Mangione(1995,第 7 章)。

与访谈相类似,调查问卷也可使用不同的形式,覆盖不同的问题。例如,问卷可以兼容封闭性问题和开放性问题。开放性的问题还可以有不同程度的"开放性":一些问题可能只要求用一个词回答,另一些问题则可能要求回答许多句话(关于如何完美设计问题的问题,见 Fowler,1993,第 5 章)。同样,问卷也可能采取

不同的形式进行提问,常见的形式是多项选择(如"请从以下选项中选择最能反映您未来的教育计划的一项"),但也可以采用清单形式的提问,即要求调查对象标出所有适于自己的选项。

人格问卷(personality questionnaires)、**人格编目**(inventories)**和人格清单**(checklists)。大多数人格清单也可视为调查问卷的一种形式,用于测量回答者的人格属性。在理论上,这些属性预计有一定的稳定性,并可以将个体相互区别开来。大多数人格属性被视为"特性",而不是"状态"。相对于"特性"而言,"状态"属性也许历时更短,且取决于具体环境。但是对人格编目进行详述已经超出了本书的论说范围,对此的总结可见 Barnett 和 Macmann(1990)。气质分析问卷(Keirsey Temperament Sorter)是人格编目的一个案例(见 Keirsey and Bates,1984)。此外,还有一些非正式但却经常运用的方法,如自我感觉、控制的位点(内部位点对外部位点)以及自我效能评价等(案例可见 Robinson, Shaver and Wrightsman,1991)。

在教育学、心理学以及其他行为科学中,有越来越多的研究者开始使用这些量表来收集与行为可能存在关联的人格属性资料。对"教师效能感"与教师参与决策之关联(Tatlor and Tashakkori,1995,1997)、"教师效能感"与接受行为干预(Hughes, et al. ,1996)之间的关系所进行的测量,就是应用这类量表的案例。自我效能感与一系列有关健康的行为如计划生育、安全的性行为以及类似行动之间的关系的问题,也是这类应用的研究案例(见 Brien,et al. ,1994;Joffe and Radins,1993)。

态度量表(attitude scales)。态度量表是通常用于问卷调查的问卷形式。创建一个正式的态度量表相当困难,而且很费时间,但既有研究中所使用过的量表数量庞大,包括对态度、信仰、自我感觉、意向、期望,以及一系列相关概念的测量(具体案例可见 Robinson and Shaver,1975;Robinson,et al. ,1991)。在其研究中使用态度量表的研究者通常会采用李克特型量表或奥斯古德型量表(Osgood Scales)(即语义差异量表)。关于这两种量表(以及其他形式的量表)的创建程序,Ary 编著的著作(1996)或 Gall 编著的著作(1996)给出了总结分析。

在李克特型量表中,要求回答者在对提题(或某种属性的存在与不存在)表示同意或不同意的程度时,应当依据由4个或5个选项(如"非常同意"、"同意"、"不同意"、"非常不同意")所组成的回答量表来表达,每个回答都被赋予特定的数值(例如,根据同意与否及其程度,赋值为+2、+1、-1、-2)。然后,研究者就可以将所有问题的分数加总起来,获得对全部问卷或其中一部分的整体态度的分值。

在奥斯古德型(语义差异)量表中,要求回答者在对某个态度指向对象(如"站点式管理"、"社会研究课"等)表达其反应时,应当在每一组由两个语义对立的形容词(如有用……无用)所构成的量表中作出标注。研究者可以对每个这样的回答赋值(如+2、+1、-1、-2),然后将这些分值加总起来以获得总体态度的分数。奥斯古德型量表的好处是,同样的一组两极形容词可以用来测量对多个态度目标的态度。例如,在问卷页顶的标题处用"数学课"替代"社会研究课"之后,还可以使用相同的形容词。

间接自述报告:投射技术。这类测量工具立基于这样的假设:在面对着一种不明朗的情势时,人们的反应体现着其感觉、信仰和倾向,但这些是在面对直接提问而给出回答时无法表现出来的(本书并不打算对这些方法进行详细介绍)。心理学的传统研究案例是墨迹法的使用(如 Holtzman 墨迹法。相关评论见 Chandler,1990),根据儿童绘画进行投射阐释则是另一个儿童心理学及相关学科的研究案例。(相关案例可见 Knoff,1990)。另外一种常见的方法是语句完形:在这一方法中,研究者为个体提供了与特定议题相关的不完整语句,然后要求个体将语句补充完整(见 Haak,1990),个体的反应随之被纳入"内容分析"之中(见本书第6章)。

看人们在做什么:观察方法

自述报告的方法尽管有其优点,但也存在着一个主要缺陷,即自述报告的测量会受到报告者自身的影响(即前面所讨论过的参与者反应性)。通常,观察法可用来直接观察人们的行为而不是询问个体对人们行为的看法,进而可减少参与者反应性的影

响。显而易见，有许多属性、动机、感受等是不能直接从行为中推导出来的。因此，观察方法如果能结合以信息收集的其他方法，将会更加有用。

参与性观察。“参与性观察”、“自然观察”和“民族志观察”往往用来指称研究者在自然情境中对行为、事件进行观察和记录的各种资料收集方法（如 Rosnow and Rosenthal，1996；Spradley，1979）。在所有这些观察方法中，研究者都需积极地参与到所观察的单位中的人际环境中，主要目标是对“自然情境”下的行为和互动模式进行测量和记录。

非参与性观察。在这种观察中，研究者并没有积极地参与到其观察的行为和互动所发生的情境之中。典型案例就是通过单向镜（one-way mirror）观察儿童的玩耍行为；另一类案例则是通过隐藏的摄像机或录音机进行记录，但这在研究中被认为是不道德的。通过记录网民对不同网站的上网模式和频度从而对拥有市场营销或其他目的的网民进行“观察”，这尽管是一个新的现象，但也是典型的研究案例。对使用这些资料以达到研究目的而涉及的道德因素，需要有严肃的思考。

就间接观察而论，最近出现一种有意思的发展趋势：使用小型电脑来监视行为、心率以及荷尔蒙的变化。例如，一些国家曾使用许多新型的小型器具进行现场检测，用以监视妇女服用避孕药的模式和频率。由此所获得的数据随后得到记录，供研究者此后下载下来用于研究。这些非直接的、非参与性的“观察”方法正日益精巧。

非参与性观察的一种特定类型是观察行为的结果，从而推断行为者的行为及其偏好。这些“非干扰性测量”（unobtrusive measures）往往是在研究对象离开了行为情境之后很久才可获得的（某些作者（如 Babbie，1996）将档案资料和文件也归为非干扰式观察的类别，但我们更乐于将之作为下文单独的一节进行讨论）。例如，研究者可以调查高校学生从图书馆借阅的图书类型。另外一个经典的案例是，通过观察某个博物馆中地毯的破损程度，以发现博物馆的哪个分区更受欢迎（见 Patton，1990）。在测量那些易受规范压力和社会愿望之影响的变量时，这种间接观察十

分有用。一般而言,这些非干扰性观察不会陷入更一般的非参与观察所面临的极其棘手的道德困境。

观察法的优劣。观察排除了对个体进行询问以了解其行为或倾向的需要,减少了受控反应的可能性,但并没有完全消除反应性(reactivity)的问题。除了上文提到的各种反应性效应以外,意识到正被观察也可能会对个体有其他的社会—心理影响。一个经典案例是 Duval 和 Wicklund(1972)所说的"客观的自我意识"(objective self-awareness)。根据他们的研究发现,当人们处于被观察的状态中时,就会变得更加注意自己的行为与价值(或所理解的规范)之间的矛盾之处,这反过来就会导致对其行为的因果关系作出不同类型的推论,还可能会导致其行为或对社会环境的反应发生无意的变化。

根据所观察到的反应和行为进而对个体的意向、目的或其他不可观察的属性作出推论是很困难的,而且常常是不可靠的。我们对社会的理解大多集中于对他人行为之因果联系的理解;而作为研究者,我们无法幸免于这种理解所具有的内生性的错误倾向。关于这种倾向的一个著名案例就是属性上的"行为者—观察者差异"(Jones and Nisbett,1972)。根据社会—心理学研究,在观察者对行动者以特定方式采取行动的原因所作出的推论判断与行动者本人对这些原因的理解之间,可能会有很大的出入。作为行动者,我们一般会将我们行为的原因归结到环境因素("外因")之上,而观察者却会将这些相同的行为归结到我们自身的特殊因素(我们的性情、信仰或"内因")之上。这类认知倾向带给我们的启示是,作为观察者,我们可能会更多地将其他人的行为视为其态度、性格、意向等的结果,而不是他们对环境进行反应的结果。

在定性研究与定量研究中,观察通常都被用来获取有关被观察者之行为的信息,而不是获取被观察者对其行为之原因的观点和(或)感受。对于这种观察,上述提及的这些认知倾向也许不能适用。而且,对于定性研究者来说,被观察者的观点是其主要的研究兴趣之所在,而非观察者或研究者自己的观点。因此,这种认知是通过访谈或其他对一手信息(如日记、文章)的自述报告进行确认和记录的。在本书第 4 章中,我们列举了确保定性研究中

资料和推论之质量的各种策略。这些策略(如成员核查)将进一步保证研究者不受上述认知错误的可能的影响。

对个体询问他与他人的关系:社会人际学

研究者可以使用社会人际学获得群体内人际互动、交流的"快照",让群体或组织内的每一个成员具体指出与自己互动的人,或者愿意在困难情况下与之互动/合作的人。这样就产生出一幅社会关系网络图(socialgram),可以表现群体成员之间的互动、相互关系的复杂网络——社会关系网络图是对这些互动和交流的图形表现。在本书第8章讨论 Durland(1996)的论文时,我们将更详细地介绍社会关系网络图的应用。

社会人际关系法不存在自述报告方法中所存在的那些反应性问题。原因在于群体内的成员不需回答关于群体结构的问题(如谁最受欢迎、谁不受欢迎);研究者向每位群体成员进行询问,确定其所互动或有机会时所互动的一两个其他群体成员。而且,群体的单个成员要分析清楚群体内人际关系演变的复杂结构,通常是很困难的。

社会关系网络图同样也具有局限性。一般说来,就群体内所存在的特定的互动模式,这一方法并不能提供有关原因的信息,这些原因大多都是通过随后的访谈或田野观察才能辨认出来的。而且,随着组织规模的扩大,社会关系网络图的复杂程度也会呈指数级增长。因此,要对大型组织和群体作出推断是极其困难的。电脑分析、制图和运算统计指标的复杂方法已经发展成形,这对研究者是十分有用的。有关这方面的详情,见 Wasserman 和 Faust(1994)以及 Durland(1996)。

使用他人收集或记录的资料:档案资料和综合分析(meta-analysis)

档案资料指的是个体和(或)机构所收集和记录的关于个体和(或)机构的资料。这种资料的典型案例有:学校关于学生的记录、诊所关于病人的记录、城市关于结婚和离婚的记录、图书馆关于某一特定事务的"口述史"收藏以及个人的自传材料。显而易见,这些资料既可以采取数据的形式也可以是有关研究对象的叙

述性信息。①

在分析档案资料以解答研究问题之前,研究者必须留心关于这些信息的质量的两个主要问题:效度和信度。另两个与大多数历史、档案信息及口述史料有关的问题,一是"资料记录得准确吗"?二是"资料保存得完整吗"?如果信息的记录不准确或只选记了一些个体/案例,研究者的研究结论的内部效度或可信性就会有问题。如果一些信息得到了保存而另一些却没有保存下来,后果同样如此。例如,如果学校的记录没有包括那些中途退学或被开除的学生的文档,相关研究结论的内部效度(可信性)就会受到质疑。

"整合分析(meta-analysis)"②所收集的信息主要是指与某一论题或问题相关的其他课题的研究发现。定量研究的整合分析是对许多研究所作的一种综合,以期在这些研究之中估计出变量间关系的力度(效应的规模),以及效应/关系在经验研究中变异的程度。在大多数应用案例中,整合分析并不被视为一种资料收集的策略,相反,它被看作一种资料分析的方法。有关详情可见 Rosenthal(1991)。

尽管人们常将整合分析看成一种定量研究方法,但这一方法的一般形式(即将此前的已有研究作为分析单位)也适用于定性研究。譬如,研究者也许并不想确定效应的规模大小,而是将一组研究发现之中的主题梳理成某种类型学或分类法。

资料收集的多元模式。毫无疑问,使用多元资料收集方法以及多元资料来源是有众多好处的。只要"研究现场的现实情况"允许,无论是定性研究还是定量研究,大多数研究都尽可能地利用这些多元方法。例如,在单一主题的实验研究(定量的)和定性

---

① 大量的信息可以从公共领域的数据库中获取。这些资料(是由国家、州甚至国际组织收集的)。对于从事"二手研究"的研究者而言,也可视为档案资料。在美国,这种数据库的例子有"社会普查(GSS)","全国八年级教育跟踪研究(the National Education Longitudinal Study of Eighth Grade)(NLS-88)"。国际组织的数据库包括联合国、世界银行和私人组织每年从不同国家收集的多种资料,以及各国政府开放使用的人口普查数据和相似的数据库。关于对这种数据的分析方法的详细讨论,见 Stewart 和 Kamins(1993)。

② meta-analysis 在指称专门用于分析和综合研究已有文献群体的方法时,亦可译为"荟萃分析"、"统合分析"或"聚合分析"。(译者注)

案例研究中(见 Yin,1994),都可以通过多种渠道来源(参与者、父母、老师等),并通过多种方法(如观察、访谈)以及标准测试,来从事资料收集工作(关于"单一案例研究"的更多信息,见 Franklin, Allison and Gorman,1996)。与此类似,实验型研究者和民族志研究者都有机会从多元的资料来源(在实验研究中,其他来源是指研究的其他参与者)之中获得多元类型的资料(观察、访谈、档案等)。甚至在大规模的调查研究中,如 1988 年的"全国教育跟踪研究"(NELS-88;Ingles,1992),研究资料都可以通过多元的资料来源(学生、父母、老师、学校)和多元的测量模式(横向调查、成绩测验、学校记录以及其他档案资料)来收集。

研究者所用的研究方法(如实验法、问卷调查法、民族志等)可能会对研究者所使用的资料收集程序之类型构成某些限制。但是,研究者可以通过多元方法和多元渠道找到资料收集的创新方式。不过这种创新只有在研究者对不同的研究程序/设计驾轻就熟之后,方有可能。有关研究方法的大多数教材都会为研究者详细介绍这些研究程序/设计。

有一类常常被忽视的资料来源,即研究者对一种文化、一个群体或一个组织的个人知识。这种知识尽管不能被系统地加以衡量,但却提供了辅助性的资料来源,可以极大地丰富研究者所收集的资料。在我们自己的研究中,我们经常会使用逸闻趣事性质的文化观察和经历,以此作为"非正式"的资料来源。在跨文化和多文化研究中,这些"文化观察"往往会导引出对通过调查或其他方法所获得的结果而进行的解释。尽管这些"专家"的观察并未纳入"研究资料"的范畴,但它们仍是定性(推论型)的研究资料。因此,我们认为,跨文化研究和多文化研究,甚至包括高度定量化的研究,其内在属性都是混合的。

我们自己正在进行的一些研究(Tashakkori, Aghajanian and Mehryar, 1996; Tashakkori and Thompson, 1991; Tashakkori, Thompson and Mehryar,1987)即可提供这样的案例。在 20 多年间(1974,1982,1984,1986 和 1996),我们从伊朗南部的青年中收集了 5 次问卷调查的资料。在每一次问卷调查中,我们都测量了伊朗青年对教育、婚姻、生育和妇女加入劳动力市场等方面的态度、意向和个人计划。我们将这 5 个时间段的测量结果根据性别和社

会经济地位(SES)的背景进行了对比。我们自己对过去30年间伊朗社会文化变迁的了解以及我们在这一时期所进行的文化观察,对我们阐释这些调查资料有非常重要的帮助。

我们的研究结果显示,在1970年代,伊朗出现了一个走向现代化的稳定趋势,但在1980年代初这一趋势又遇剧挫。然而,与我们一些同事的预测相反,研究结果同样显示出,在1990年代又出现了一股回归现代态度和行为的较强趋势,在妇女当中尤其如此。如果没有细致的文化观察和知识,对这些看上去并不一致的趋势和变迁进行解释将是不可能的,或至少是不准确的。对于一个混合模型的研究者来说,即便推论是建立在高度系统化的定量资料的基础之上,但对资料收集之场景的个人观察,以及与提供资料的个体进行互动,等等,都是非常宝贵的信息来源。甚至在高度结构化的实验研究中,这样的观察对基于研究结果而作出的推论的质量也是大有裨益的。

6

传统资料分析策略的替代性选择

我们将在本章首先就定性研究传统和定量研究传统中的资料分析策略给出一个概览。其间，我们将分别讨论这两种一般方法的优劣长短。然后，我们将说明在同一项研究中将这两种分析类型结合起来的替代性方法，并列举来自学术文献中的研究案例。

正如我们在本书第4章所讨论过的那样，资料的分析和说明策略取决于研究变量的测量量表。就名义变量和定性描述型(qualitative-narrative)变量而论，宜采用定性的资料分析方法；而大多数顺序层次(ordinal-level)的变量，应采用非参数统计(nonparametric statistic)作为资料分析方法；参数统计主要运用于等距层次(interval-level)的变量。尽管存在着这些不同，参数数值分析同样经常运用于理论上的等距变量，即便它们所获得的测量结果可能是顺序性的。

传统的定量资料分析

大多数传统的资料分析方法可以用一个由两个维度交叉所构成的2×2定量资料分析矩阵作出概括。其中的两个维度是：①是否使用统计工具来判断变量间的关系或组群间的差异；②统计工具是否适用于等距或非等距数据。对于大多数研究新手来说，这一简化表达自有其好处，但老练的研究者可能会有他们自己归类总结统计程序的方式。我们在此对这一简化表作出介绍，以便引出将在本章的后面详细说明的统计程序的不同“家族”。

尽管这个 2 ×2 矩阵可以很好地概括不同"家族"的统计程序,但本章对定量分析方法的其他讨论是集中在另外的两个二分法之上,即:

1. 描述性方法对推断性方法
2. 单变量方法对多变量方法

描述性方法

描述性方法包括通过简单的统计和图表展示来呈现研究结果。这些分析的主要目的是提供便于读者理解变量性质及变量间关系的图形和(或)总结。最常用的描述性资料分析和说明的方法是:①集中趋势的测量;②相对地位的测量;③变量间联系/关系的测量。

集中趋势的测量是将一组观察点或数值总结为一个数值。众数、平均值和中位值都是集中趋势的测量手段,都是呈现构成组群的事件或人的单一数值:众数(mode)是一个组群之中以最高频度出现的那个数值,平均数(mean)是指各个数值的平均数(即各个数值的总和除以数值的个数),中位数(median)则是指 50% 的数值都落在这一数值以下(中位值将数值组分为两个等半部分)。

那么,对于集中趋势,哪一种测量手段更好呢? 对这一问题的回答要取决于测量量表的类型。众数是一个通俗易懂的简单测量手段,但它只代表组群中的一个数值,不能提供关于其他数值的信息。

平均数是组群分布的所有数值的平均,因此它包含了组群内所有成员的信息。但是,在这一长处之外,它也存在两个缺陷:第一,它只能由顺序量表或更高要求的量表所测量的变量计算出来,而如女性群体和男性群体的性别"平均值"则是没有意义的;第二,极端数值会扭曲平均值。

中位数是指处于分布于正中间的数值。例如,如果参加某匿名聚会(alcoholics anonymous,A. A.)的一组成年人的年龄分别是 17、19、20、56 和 59,那么,这组人的中值年龄就是 20。这一数字与平均值年龄(34.2)相比非常不同,呈现出的显然是该组人群的不同形象。如果在组群中加入两个新成员,一个是 16 岁,一个是 85

岁，中值也不会因此改变。这是因为中值不包含组群中其他数值的信息。在某些案例中这是一个缺点，但是，在那些分布中存在着极端数值的案例中，如有85岁的匿名酒会成员时，中位数就具有优势。

变异性测量(measures of variability)包括平均偏差、方差、标准偏差和四分位数间距。平均偏差、方差和标准偏差都立基于每一个数值与平均值之间差异的单位数量。它们代表的是各个数值在分布中集聚于平均值周围的程度(相关评论可见Gravetter and Wallnan,1998)。在不考虑偏差的正、负值的时候，平均偏差是指所有偏差简单求平均所得的值(即平均偏差是指各偏差的绝对值的平均)。

如果将每个偏差平方后相加，并将所得的偏差平方和除以数值的个数，那么，研究者就会得到方差(即方差是偏差平方的平均值)。从概念上说，方差的平方根与平均偏差相似，我们称之为标准偏差。方差的使用比平均偏差更为频繁，因为它比平均偏差更易于进行数学操作。

相对地位的测量(measures of relative standing)只表示某一数值对于其他数值的相对位置。百分等级(percentile rank)是其中之一。百分等级只简单地显示有多少百分比的数值位于某一数值之下。例如，如果某学生的数学分数在“加利福尼亚成绩测试”中的百分等级是68，这就意味着该学生的成绩比参与测试的组群中68%的测试学生的成绩要好。百分等级简单明了，但仍有一个大的缺陷：等级间的差异并不代表能力或潜在属性之间的差异。例如，某一特定的两个百分点等级差(如70－68＝2)往往并不等于其他一些两个百分点等级差(如88－86＝2)。

标准数值则避免了这一缺陷。一个案例就是*z*-值。无论原始数值如何，两个*z*-值的差都有着相同的意义(例如，某一*z*-值的两个点差与其他*z*-值的两个点的差，如2.8－0.8＝2，所表示的属性数量的差距是一样的)。若某人的*z*-值是建立在其数值与分布的平均值之间的偏差之上，则将该偏差除以分布的平均偏差，即获得z-值(也就是说，将某人的偏差值转化为若干单位的标准偏差)。

联系/关系的测量(measures of association/relationship)只表示两个变量或更多变量之间的关系的程度。Pearson相关系数(*r*)

是应用最广泛的范例。Pearson 相关系数(r)处在从 -1(完全否定关系)到 +1(完全肯定关系)之间。相关系数接近 0 则意味着没有关系。如表 6.1 所示,在相关系数的家族中,还有大量的其他统计方式(如 rho,phi;见 Gravetter and Wallnan,1998)。

表 6.1 定量资料分析矩阵

社会和行为科学中使用的定量资料分析方法的简单分类

数据类型	变量间关系	组群间差异
间距/顺序	Pearson 相关(r)	独立样本的 t-检测
	多元相关	ANOVA/ANCOVA
	标准相关	MANOVA/ MANCOVA
	回归分析	区别性分析
	因素分析	
顺序/名义	Rho	信号测试(sign test)
	无关/相关的卡方检验	Wilcoxon 配对
	Phi	
	Cramer's V	
	逻辑回归	

推断性方法

对于大多数研究目的而言,描述型统计、图表以及二者的结合还是不够的。更具体地说,这些方法尚不足以完成评估和假说检测。用于检测假说的资料分析方法,其基础是对确定组群间差异或变量间关系的过程中有多少误差所进行的估量。这些资料分析的方法通常被归类为推断性统计,典型之一就是检测两组平均值间差异之显著性的 t-检测。下面将对此进行详细论述。

在推断性统计分析中,对于结果的发生究竟有多少可能出于"单纯的偶然和随机误差",还是源自变量间某种真实的基本关系,统计显著性的检测可以提供相关信息。如果检测的结果(如平均值之间的差异)具有统计学意义上的显著性,那么,研究者就可以得出结论,结果不仅仅是偶然发生的。这种假说检测所包含的基本假设是,任何变量间的表面关系(或组群间差异)事实上都

有可能是由变量测量或被观察/被研究个体的随机变动引起的，而推断性统计就是评估这种偶然变动的程度的方法。此外，这种资料分析方法还可提供有关结果的量级或关系的信息。这些方法可归纳总结如下：

1. 检测群组平均值间的差异：
   a. 比较样本的平均值与总体的平均值：*z*-检测。
   b. 比较两个样本的平均值：
      ①独立观察组：对独立组的 *t*-检测；
      ②非独立观察组（配对组、重复观察等）：对非独立组的 *t*-检测。
   c. 比较两个样本或更多样本的平均值，或比较因子设计（它们有一个以上的自变量）的平均值：方差分析（ANOVA）。
   d. 在对外部变量引起的变化进行控制的同时，比较两个或更多样本的平均值：协方差分析（ANCOVA）。
2. 确定相关系数（或回归斜率）是否真不为 0：
   a. 对从 0 开始的 Pearson（*r*）显著性进行 *t*-检测。
   b. 对多元相关显著性进行 *F*-检测。
   c. 对多元回归分析斜率的显著性进行 *t*-检测或 *F*-检测。

单一变量方法对多元变量方法

在上一节，我们所列举的案例和进行的讨论都关乎于单一变量的统计学。这些统计分析立基于一个（因）变量，或是在双变量相关的情况下，显示了两个单一变量之间的关系。但社会和行为科学的许多研究课题都会寻求两个以上的单一变量间的关系。例如，假设研究者想要发现某一变量（如大学平均成绩 GPA）与另一些变量（如高中的平均成绩、父母教育、父母收入）的综合影响之间的相关性，那么在这个大学平均成绩的案例中，便需要使用多元相关分析（*R*）。多元相关分析仍然是一种单变统计分析，其解释非常类似于双变量的（pearson）相关分析。

在多元相关（或多元回归）分析中，等式的一边是一组变量，另一边是一个单变量。但有的时候，研究者却有意在等式两边将多元变量联系起来。这样，当我们在研究两套（组合）变量间的相关性时，典型相关（canonical correlation）就会是可选择的统计分析

方法。例如,研究者也许会有意探讨自我效能感、学术自我观、去年的平均成绩三者与高中学生三种有关成绩的指标(数学课、科学课和阅读课)之间的典型相关性。在这里,典型相关分析是一种多变量的统计分析。

其他多元变量方法的典范有区别式函数分析和要素分析。在区别式函数分析中,研究者的目的在于发现可以将两个组群或多个组群彼此区分开来的一套最佳变量。譬如,什么变量可以对辍学的八年级学生与继续升读到十年级的八年级学生进行区分?

在探索性的要素分析中,研究者的目的是要确定一组测量/变量的潜在维度(概念)。在确证型要素分析(与大量的变量或要素分析的延伸如结构性方程模型)中,研究者的目的是确定概念的预测结构是否能从资料中得到(关于多元变量方法的更多信息,见 Syevens,1996,and Harris,1985)。

传统的定性资料分析

与定量资料分析一样,定性资料的分析也有许多既定的程序。要对这些叙述性的资料进行分析,就要有事先的准备,通常先须将原始材料(如田野笔记、文件、录音带)转化为半成品的资料(如补充、登录),然后将之编码,纳入到特定的分析框架之中(Huberman and Miles,1994)。

这类分析框架可以依据两个维度加以区分:①研究框架或范畴是分析之前确定的,还是产生于分析过程之中;②定性分析框架的复杂程度(从简单到复杂)。表 6.2 列出了由这两个维度交叉而构成的一个定性资料分析矩阵,并附带了一些关于定性资料分析之具体策略的案例。

在介绍这些定性资料分析的具体策略之前,我们要对构成定性资料分析矩阵的两个维度作些说明。这一矩阵还包含了一些研究框架的案例,它们已经被众多分析人员用来说明叙述性资料的含义。这些分析人员处理定性资料的方法有许多不同的取向,包括后实证主义、实用主义以及建构主义。

例如,Huberman 和 Miles 在《定性研究手册》一书关于资料处

理和分析的一章中，将自己称为“先验的现实主义者”，他们“相信社会现象不仅存在于思想之中，也同时存在于客观世界之中，而且它们之间存在着某些有规律的、合乎理性的、稳定的关系”(Huberman and Miles,1994:429)。Huberman 和 Miles 还进一步指出，定性资料既可以采取归纳方式进行分析，也可以采取演绎方式进行分析，这取决于研究者正处于研究之轮的哪个阶段。对于为评估教育创新而收集的定性资料，他们在分析的框架内事先建构了他们所指称的“效应矩阵”。这一矩阵所依据的是他们以前在分析同类资料时所获得的经验(Huberman and Miles,1994)。

这种观点与建构主义者的主张恰好相反。在建构主义者看来，定性资料分析过程中应当只使用归纳逻辑，同时，分析框架也应当从资料之中生成，而并不能先验地加以确定(如见 Lincoln and Guba,1985)。例如，Lincoln 和 Guba(1985)提出了一个扩充版的常数比较法(constant comparative method)作为他们分析定性资料的方法，这种方法最早是由 Glaser 和 Strauss(1967)提出讨论的。这一方法体系应用归纳逻辑以逐渐从大量叙述性资料中形成分析的框架或范畴。Lincoln 和 Guba 两位很可能会认为，Huberman 和 Miles(1994)所先验设计的矩阵并不适于分析定性资料，因为这些矩阵立基于经由演绎而推出的范畴。

表 6.2 定性资料分析矩阵

对社会和行为科学中定性资料分析方法的简单分类

框架类型	简单框架	复杂框架
先验的框架 (a priori)	简单价(valence)分析 显性内容分析	效应矩阵 (Huberman and Miles,1994)
生成的框架 (emerging)	隐性内容分析 常数对比分析 (Glaser and Strauss,1967; Lincoln and Guba,1985)	发展型研究顺序 (Spradley,1979,1980)

就复杂性维度而论，定性资料分析矩阵的多种不同框架之间构成了一个连续谱。最为基础性的框架是简单价分析，其案例是 Teddlie，Kirby 和 Stringfield 所作的研究(1989)：在他们的分析中，

有三个预先确定的范畴用于对教室观察所收集的定性资料进行编码。更为复杂的分析框架则是 Spradley(1979,1980)所描述的发展型研究顺序:这种生成性主题的框架包括有 12 个步骤和 3 个复杂性依次递进的层次分析(领域分析、类型分析、构成分析)。

重点在于提炼范畴或主题

任何类型的定性资料分析的本质,都在于形成一个能对众多叙述性资料进行概括的范畴或主题的类型学。正如 Huberman 和 Miles(1994:431)指出的:

> 定性研究的最终目的是对某种样式的关系进行描述和(在某种程度上)作出解释,而这只能通过一套概念意义上特指的分析性范畴来实现(Mishler,1990)。从这些范畴出发(演绎),或逐渐得出这些范畴(归纳),都是合理有效的方式。

有意思的是,在许多进行生成性主题分析的研究案例中,所形成范畴的数量似乎存在着一定的规律。通常,在对与某个特定的开放性问题相关的 200 个或更少的信息单位进行数据库分析时,大约会形成 6 ~ 8 个类型/范畴。这样的数据库最初形成的范畴数目可能是 10 ~ 15,但通过合并相似的回答组以及提出更宽泛的概念内涵,范畴的数目可以降至 6 ~ 8 个。这种分析现象的产生在一定程度上可以归结为我们人类作为生物物种在理解任一复杂主题时,其能力是有限的。

先验性主题分析的案例

表 6.2 给出了 3 个先验性主题分析的案例:简单价分析、显性内容分析以及 Huberman 和 Miles(1994)的效应矩阵。

简单价分析

简单价分析是表 6.2 列出的所有类型的主题分析中最为简单的一种。Teddlie 等人(1989)曾在一项利用某个大型数据库进行分析研究的过程中使用过这种简单价分析。这一数据库容纳了对学校的效能进行跟踪性课堂观察所获得的大量资料。在相关研究中,研究者使用带有 15 个表达教育效能一般性指标项目的开

放性工具,对700多个课堂进行观察。于是,在这一数据库中产生出了超过1 000个开放性的回答有待分析。为了加快分析进程,研究者使用了特定的编码框架,由两位评估者对每个回答样本进行分析,并将每个回答编码到3个预先确定的范畴之中。这就是:

1. 对这一特定的教学成分,所作的回答可以证明教学行为的有效;
2. 对这一特定的教学成分,所作的回答可以证明教学行为的矛盾性;
3. 对这一特定的教学成分,所作的回答可以证明教学行为的无效。

Patton(1990)将这种编码框架称为"分析家构建的类型学"。在本章的后面部分,我们在讨论所谓"定量化"(Huberman and Miles,1994)的混合分析工具时,还会对这一研究中的数据进行讨论。

显性内容分析

内容分析有两种形式:显性内容分析和隐性内容分析。Berelson(1952:18)将显性内容分析界定为"对传播的显性内容进行客观、系统、定量描述的研究技术"。在1950、1960年代实证主义取向的社会学和大众传播学研究领域里,这种类型的内容分析被狭义地界定为"将标准化的测量应用于计量上被确定的单位,并用于特性表征和比较研究文件的一种定量取向的方法"。(Manning and Cullum-Swan,1994:464)。相关的研究案例有对教材、大众杂志和报纸、经典作家(如莎士比亚)著作以及政治演说等的分析和比较。

用于这类比较分析的编码程序往往得到了尽可能的高度标准化处理。Gall等人(1996:525-526)曾这样描述这一程序:

> 一旦使用合适的抽样方法选择了内容,就需要发展出一个编码或分类体系对内容进行分析。可能的话,可以使用在前人研究中已经提出的编码体系。一方面,这一选择可以节省用于提出自有编码体系的时间,因为对于大多数内容分析的研究课题而言,建构自己的编码体系是一件既费力又耗时

的任务;另一方面,使用标准的编码范畴还可以将自己的研究与其他使用相同体系的研究进行比较。"

内容分析词典的一个典范是应用于心理学和社会学领域的《哈佛Ⅲ心理学词典》。这个词典包括了诸如文化过程、社会情感行为、角色和心理过程等领域的内容范畴。

但对于许多研究项目而言,并没有这种高度结构化的内容分析词典可资利用。事实上,内容分析在1990年代的讨论中更多地被看成是一种可对任何类型的叙述性资料进行分析的一般性方法。Patton(1990)将内容分析称为一般性的定性资料分析,并将之与定量数据的统计分析进行对比。社会和行为科学的研究者往往必须在研究之前先验地形成自己的编码框架,或在其研究过程之中逐渐形成自己的编码框架。

文件的显性内容与隐性内容的不同,是指文章的表面含意与其隐含含意之间的区别。例如,人们可以对某一电视节目中发生的(预先界定好的)暴力行为进行计数,并对节目中所展示的显性暴力的程度作出结论,但是,要真正弄清这一特定节目中有关暴力的隐性内容,还必须对节目的"语境"进行分析(Manning and Cullum-Swan,1994)。在这个研究案例中,特定的语境是这一节目的描述或情节。电视节目可能有许多暴力镜头,但其隐含的主题是角色之间的信任或关爱,这就会产生出一个非常不同于显性内容分析的隐性内容分析。

表6.2所列出的定性资料分析矩阵之中,显性内容分析被归类为一种预先确定主题的研究过程,而隐性内容分析则是一种生成性主题的研究过程。例如,(美国的)高中历史课本的显性内容的范畴可以预先加以确定,可能涉及民主理想、西部扩张、奴隶制或天定命运(manifest destiny)等。另一方面,这些高中历史课本的隐性内容在未对资料分析之前,是不能预先确定的。我们将在下一节列举隐性内容分析的更多案例。

Huberman和Miles的效应矩阵

Huberman和Miles(1994)在《定性资料分析》一书中强调,必须将定性资料放入资料收集之前或资料收集早期阶段中所形成的矩阵之中。这种先验的编码和矩阵是由这两位作者在教育创

新领域内花费了几年的工夫所形成的。他们的先验编码框架包括诸如创新的属性、创新的外部背景、创新的内部背景、创新的采纳过程、创新点演变及其转型。他们发展出了一个相当广泛的编码框架，用以将他们所收集的每一份叙述性资料进行归类处理。例如，其中一个编码是“内部环境：规范和权威”，其首字母缩写词是 IC：NORM，所以纳入这一范畴的每一个信息单位便被编码为 IC：NORM。

他们的矩阵由一个不同层级的编码范畴与另一个有不同层级的编码范畴交叉构成。例如，将援助位置与援助类型交叉起来便可构成一个矩阵。

Huberman 和 Miles（1994）认为，定性资料分析有 3 个部分：①资料还原，即在得到原始资料后，采用上面所提及的编码对原始数据进行简化和转化；②资料展示，即通过整理有序的信息将资料展示出来，由此还可以得出结论（效应矩阵便是在这一分析阶段中加以使用的）；③得出结论和论证，即确认所有资料的含义，并确定研究结论的效度。

生成性主题分析的案例

表 6.2 中有三个关于生成性主题分析的案例：隐性内容分析、常数比较分析、发展型研究顺序。

隐性内容分析

正如前文所提到的，显性内容是指文本的表面含义，而隐性内容则是指文本所叙述的潜在含义。文本的隐性内容取决于对叙述的整体内容所作的主观评价。在前文所提及的案例里，某电视节目中暴力的显性内容可以通过简单地计算节目中发生暴力行为的次数而轻松地加以确定。但在进行资料收集之前，暴力行为应当有非常精确的界定——这一界定将用来对许多不同的电视节目进行判读。

对电视节目的隐性内容进行确认，自然会困难得多。在大多数情况下，对隐性内容进行确认不可能在一个简单的先验编码框架基础上展开。对电视节目的情节必须进行仔细地检查，然后才可能生成适用于这一特定节目的分析框架。例如，叙述一伙匪徒

绑架一名年轻女性的电视节目,根据显性内容分析可能会得到非常高的暴力评分,因为节目过程中的暴力威胁是很明显的。另一方面,假如那名年轻女性与绑架者之一发展出了一种积极的、同情性的关系,这个绑匪在节目的最后还释放了她,那么,较之于其显性内容所表现出的情形,这个节目的隐性内容在暴力性程度上将会低得多。与整个节目有关的主题分析框架将会从对节目本身的分析中逐渐形成,而不能先验地确定;尽管这个框架一经确定,就有可能随即应于对其他节目的隐性内容进行分析,但这仍未改变生成性的特征。

常数对比分析

常数对比分析框架最先由 Glaser 和 Strauss(1967)提出,Lincoln 和 Guba(1985)随后对之进行了完善。这种分析框架包含两个一般性的过程:①单位化,即将文本分割为不同的信息单位,这将成为确定范畴的基础;②范畴化,即将那些关乎同一内容的单位全部归入临时的范畴之中,提出可以描述范畴之属性的规则,使每组范畴内在自洽且整套范畴互不相容。整个范畴化过程包括 10 个步骤,其中一些是重复性的步骤(Lincoln and Guba, 1985:347-351)。

发展型研究顺序

James Spradley 的发展型研究顺序是最为复杂的一种框架,用以确定与他所谓的"文化情景(cultural scene)"相关的主题。对访谈和观察资料都能进行分析的、由 12 个步骤构成的过程中,包含 3 个阶段的资料收集和 3 个阶段的资料分析(Spradley,1979):

1. 首先是提出宽泛的描述性问题(descriptive questions),Spradley 将这些问题称为综览性(grand-tour)或细览性(mini-tour)的问题;

2. 运用题域分析(domain analysis)对因回答这些描述性问题而生成的数据库进行分析。文化题域由三个部分组成:表面术语(the cover term)(文化题域的名称)、内含术语(the included term)以及语义关系(semantic relationship)(内含术语与表面术语之间的联系)。语义关系的案例之一是"严格包含"(strict inclusion,即

类属于)。例如,大一新生、大二学生、大三学生、大四学生都类属于大学生。

3. 接下来是提出结构性问题。这些问题是向受访者询问他们的文化知识是如何组成的,换言之,向受访者询问其文化知识的基本题域或单位。这些问题是用来对从题域分析中发展出来的临时题域进行确证或否证,并为已确证的题域引出更多的术语。

4. 运用分类法分析(taxonomic analysis)对因回答结构性问题而生成的数据库进行分析,进而展现出某一题域内所包含的全部术语间的相互关系。这些分类法通常包含几个分支,它们由所含术语分叉而形成的一套次级术语所组成。

5. 民族志研究者使用对比性问题(contrast questions)以发现他们用来对其所在世界中的事件和事物进行区分的意义维度。这些问题依据对比原则,意在发现某个题域(或分类)中包含的术语如何有别于另一个题域(或分类)中包含的术语。这些问题包括若干种类型,如对比确证问题、二元对比问题和三元对比问题。

6. 运用构成分析(componential analysis)对因回答这些对比性问题而生成的数据库进行分析。这一过程涉及对与文化情景相关的属性或意义因子进行系统的考察。Spradley(1979)认为,当分析人员发现某一类型内的成员存在反差时,这些反差可以看成这一类型的属性或意义因子。例如,Spradley 曾从事过一项有关监狱室友的研究,他发现"流动性"(指室友在监狱内、监狱大院和监狱外部自由出入的能力)是将监狱人群内不同类型的同室者区分开来的一种属性。由此,构成分析也揭示出流动性(或移动自由)是与监狱生活相关的主题之一。

相似与相异原则的应用

Spradley(1979:157)明确地界定了定性资料分析中使用的两个主要原则:相似原则和相异原则。相似原则认为,一个符号的意义可以通过发现它与其他符号如何相似而得到揭示;相异原则主张,一个符号的意义可以通过发现它与其他符号如何相异而得到揭示。

相似原则其实在上述所有分析体系中都得到了应用。例如,范畴是通过寻找有着相似内容的信息单位(常数对比分析),或通

过寻找有某种相似含义的术语(发展型研究顺序)而得到确定的。如果说确定某种能够概括大量叙述性资料的范畴类型学或主题类型学是定性分析的要义所在,那么,相似原则就是这一分析过程的导引,因为相似原则使分析者得以探求资料中的共同之处。

相异原则在发展型研究顺序中的构成分析阶段得到了明确的使用。它也是常数对比分析的一部分,因为它指导着研究人员探求所提出的各种范畴类型之间的互斥性。

混合方法研究的替代性分析策略

混合分析方法研究的替代性分析策略的主要目的,是使研究者能够在同一项研究中同时使用或先后使用定性、定量两种传统类型的分析路径。

在社会和行为科学领域,研究者已经开始广泛地采用许多不同类型的混合方法。框 6.1 展列了 Caracelli 和 Greene(1993)对这些方法的总结。

框 6.1 Caracelli 和 Greene 关于混合资料分析策略的总结

(1)资料转化:将一种类型的资料转换(或转变)为另一种类型的资料,以便对两种类型的资料同时展开分析;

(2)类型学提炼:对一种类型的资料进行分析而产生某种分类法(或一组实质性范畴类型),然后将这一分类法用做对相对的另一种类型的资料展开分析的框架;

(3)极端案例分析:"极端案例"可以从一种类型的资料分析得到确认,并通过(补充性资料收集以及)另一种类型的资料分析进行追问,旨在对前期的极端案例之解释进行检测和完善;

(4)资料协同/资料合并:对两种类型的资料进行联合分析,建构出新的变量或协同性的变量和资料库,使之既能以定量方式进行表达,亦能以定性方式进行表达。这些协同性的变量或资料库随之会在进一步的分析中得到使用。

来源:Caracelli 和 Greene(1993)。

在此,我们将以有所不同的逻辑对混合资料分析的策略进行分类。表 6.3 展列了这些策略分类的概要。本章的其他部分将对

这些替代性的资料分析方法展开讨论和案例说明。

混合方法中一个主要的资料分析策略，是将采用定性或定量中的一种方法所收集的资料转化为另一种资料，以便利用替代性的方法来分析同一资料。这种转化包括两个方面：①将定性信息转化为可以作统计分析的数字代码；②将定量数据转化为可以作定性分析的叙述。我们将第一种转化称为“定量化方法”，而转化后的资料则是“定量化资料”；第二种转化则称为“定性化方法”，由此转化后的资料称之为“定性化资料”。

对（定性和定量）两种类型的资料进行平行分析，可以为理解变量及其相互关系提供更丰富的信息。但是，它也使研究者受到囿限，对一个子类的资料只能用一种类型的分析方法（定性或定量）。而在许多研究中，研究者可以通过下述各种方法从资料中获取更多的洞见：

1. 对同一资料同时进行定性的资料分析和定量的资料分析；

2. 对同一资料采用不同的方法（如定量）进行二次分析，以确认或扩展从前一种资料分析方法（如定性）所得出的推论；

3. 先用一种分析方法获得研究结果（如通过定性分析将个体分类成不同的组群），然后以此作为起点，用另一方法对其他资料进行分析（如用统计方法比较研究用定性观察所发现的不同组群）；

4. 以某一分析路径的研究结果（如初步访谈或文本内容分析）作为起点，采用另一分析路径来设计进一步研究的步骤（如工具开发），或收集新的数据。例如，许多调查问卷都是在对适当的人群进行了初步的定性研究之后而设计出来的。

表 6.3　替代性混合方法资料分析策略的分类

(1)同步型(concurrent)混合分析：同时进行定性资料分析和定量资料分析。
(a)不同资料的同步分析：平行混合分析（也称资料来源的三角测量）
(b)同一资料的同步分析：定量化
(c)同一资料的同步分析：定性化
(2)顺序型定性—定量分析：先进行定性资料分析，随后通过定量数据的收集和分析进行确证。

续表

(a)通过定性资料分析/观察而建构不同的人或环境的组群,然后运用定量数据对这些组群进行比较研究(如 MANOVA,整群分析(cluster analysis),区别式方程分析(discriminant function analysis))
(b)通过定性分析(如内容分析)而建构不同的属性/主题的组群,随后进行确证性定量分析(如因子分析、结构等式建模(structural equations modeling))
(c)通过探索型定性分析建立某种关系/因果联系的理论顺序,然后通过定量数据和分析(如路径分析、结构等式建模)而对前述理论顺序进行确证*
(3)顺序型定量—定性分析:先进行定量资料分析,随后进行定性资料的收集和分析。
(a)在定量数据基础上建构不同的人群或环境的组群(如整群分析),然后通过定性资料进行组群间比较研究*
(b)通过探索型定量分析而建构不同的属性或主题的组群(如因子分析、多维度量表),然后通过新获得的(或新的)定性资料和分析进行确认(如常数对比法)
(c)通过探索型定量分析(如路径分析、结构等式建模)建立某种关系/因果联系的理论顺序,然后通过定性资料和分析进行确认(如观察和个体访谈)*

* 这两种策略需要进一步完善,在此我们仅将之列置出来,不作进一步讨论

尽管我们认为上面所开列的策略清单并不全面,但它已经覆盖了社会和行为科学领域内大量的研究案例。下面几节将为这些混合策略列举相应的研究案例。框 6.1 中也列出了对这些策略所进行的分类,但与我们的分类略有不同(Caracelli and Greene,1993)。

同步型(Concurrent)混合资料分析

1. 平行的混合分析。定性资料和定量资料的平行分析,也被称为资料来源的三角测量法,大概是在社会和行为科学领域内应用最为广泛的混合资料分析策略。许多研究者在其研究中兼容并蓄地收集了定性资料和定量资料。在实验室实验中,研究者往往会在最后对实验的参与者进行访谈(后实验访谈),以确定可能影响到其回

答的个人理解和观念类型，而在实验中对参与者的观察则是又一种资料的来源。研究者在对所获定量数据进行统计分析的同时，再通过内容分析对访谈资料进行（或可以进行）分析。

框 6.2 一个对定性资料进行统计分析的案例

Roberts 和 Le-Dorze（1994）在研究失语症病人的表达流畅性时，使用定性方法收集了相关研究资料。但他们对所收集到的资料同时运用了定性方法和定量方法进行分析。这项研究的目的是发现前人研究中没有发现的失语症病人所具有的表达流畅性。研究资料是通过访谈两组患有失语症的病人而进行收集的，一组是初患者，另一组是长期患者。访谈中包括一项对表达流畅性的测试，其要求病人尽可能地说出属于同一范畴的不同词汇。研究者对每名患者在间隔 6 ~ 8 周的时间内进行了两次访谈。定量资料分析包括使用方差分析（analysis of variance，ANOVA），以群组和访问数量作为自变量；正确得分的总数作为因变量；定性分析则包括将错误归类，确认患者回答里所存在的语义联系及形式联系。随后，研究者对这些资料进行了定量化转换，以便能进行统计分析。定性分析显示出，不同的病人群组对结果有显著的影响，因为初患组说出的正确词语更少；同时，访问次数也对结果有显著的影响，因为两组患者都在第二次访问时表现得更好。但是没有发现群组与访问次数之间相互作用的影响。对定量化数据的分析显示出两个有意思的结果：第一，与此前的研究不同，两组患者在正确词语的数量与错误词语的数量之间都呈现出显著的负相关；第二，患者说出的有语义联系的词语数量与正确词语的总数量之间存在着显著的相关性。

在问卷调查的研究课题中，经常会有开放式的回答选项和封闭式的回答选项并列一处的情形（见本书第 4 章）。研究者对封闭式的回答进行统计分析，又对开放式回答进行内容分析。在高度非结构化的定性研究和田野研究（见 Babbie，1966）中，尽管大批资料是定性的，研究者也会相应地进行定性分析，但也有一些资料是（或可以是）进行定量分析的。这种定量分析的最简单的形式，就是对适当的变量进行描述性统计性质的计算（具体案例可见 Gall，et al.，1996：168-171）。

在大多数其他类型的研究中，也都可以找到与平行资料收

集/分析相似的类型。这是教育学许多研究课题所具有的特征。在教育学研究中,定量资料(如测验、对教师课堂行为的正式测量)是与定性资料(如非正式的学校观察或对同事、教职工的访谈)同步得到收集和分析的。

2. 两种方法应用于同一种定性资料的同步分析,包括将定性资料转化为数值形式。此前,我们将这种转化称为“定性资料的定量化”。一方面,定量化可以包括对特定主题、回应、行为或事件进行简单的频点统计;另一方面,定量化还可以包含对这些事件、行为和表达的强度和力度进行更为复杂的定级评估。借助于这种类型的转化,不同的定量方法可在分析中得到应用。譬如,描述型统计会被用来进行总结或框定频点统计,诸如因子分析、相关分析或回归统计等更为复杂的分析程序则可以在定级评估的基础上展开。框 6.2 给出了这种转化类型的一个案例,本书第 8 章所提到的 Rusbult、Oniznka 和 Lipkns(1993)的研究则是另一个案例。

3. 两种方法应用于同一种定量数据的同步分析,包括将定量数据转化为定性的范畴或叙述。此前,我们将这种转化称为“定量数据的定性化”。这种转化的一个典型研究案例是 Hooper(1994)对多元文化课堂中的语言艺术作业之效果的研究(见框6.3)。

定性化工具还包括叙述性形象的构建(narrative profile formation)。在定性化过程中,调查者对受调查的个体形成文字描述。在这里,至少有 5 种相互关联的形象类型可以从定量(或定量与定性的综合)资料中建立起来。这些形象的类型(情态的、平均的、整体的、比较的、规范的)并不是彼此间相互排斥的(也就是说,描述的结果在某些方面可能是重叠的)。因此,叙述性形象至少在某些方面与类型学的构建是相似的,后者将在下文的另一节中讨论。值得注意的是,这里的形描是同步分析的结果,而类型学则是在研究者进入顺序分析时所形成的。

关于形描的 5 种类型有如下表述:

情态式形描。这是根据群体最常见的属性对某个群体(如妇女群体)所进行的详细的叙述性描绘。例如,如果大多数的个体都是 50 岁,那么这个组群就可以被视作中年人。

框 6.3　关于定量数据定性化的一个案例

Hooper(1994)曾经进行过这样一项研究，其涉及读写语言艺术的简单任务和复杂任务对多元能力、多元文化课堂上的学生之动机和学业的影响。尽管立基于前人的研究和相关理论之上，但这一研究仍具有开创性。研究的主要目的是考察何种类型的任务可以激发学生的内在动机。Hooper 使用了定量和定性两种方法来选择作为研究对象的课堂。课堂的选择标准有 3 个：能力范围、任务的整体分布以及语言艺术和阅读主题的范围。在挑选出课堂组群的基础上，又随机地选择出 3 个组群的学生：从 3、4 年级选出的一般能力的学生，从 6 年级选出的多元能力的学生，以及从 3—6 年级选出的缺乏学习能力的学生。

资料的收集工作同时结合使用了定性方法和定量方法。Hooper 对每位学生都进行了访谈，询问他们对两个简单任务和两个复杂任务的理解、期望和信念。一些问题要求口头回答，由此构成了定性资料；另一些问题则要使用评估量表对问题进行回答，由此构成了定量资料。研究者通过对每个分数段划分出一个类型，将定量资料转化为定性资料。分析工作在组群层面和个体层面同时展开。在组群层面上，分析工作包括通过将学生的兴趣水平与动机聚焦水平交叉起来而进行编码；在个体层面上，分析工作包括将任务类型、语言回答和学习理由交叉起来。Hooper 的研究发现，当语言艺术和读写任务从简单变为复杂时，学生的动机也趋向于内在理由。另外，这项研究也发现了这样一种可能性：依据年级、性别、种族、社会经济地位(SES)和能力而区分的组别可能不如以前想象的那样影响学生的动机。

平均式形描。这是根据个体、情境的许多属性的平均状态(如平均值)进行叙述性的形描。这种形描是在这些平均属性的基础上对组群进行详细的叙述性描绘。

整体式(推论性的、概括性的)形描。这种叙述性形描是由调查者对被调查单位的整体印象所构成。与平均式形描不同，这种整体印象所依据的特定信息也许是无法明确表达或直接获得的。

如前所述，上述这些不同的形描类型可以混合使用。例如，

可以根据组群的最常见属性、某些平均属性以及最终的总体(整体)推论来建构某一种形描。Tashakkori, Boyd 和 Sines(1996)的研究可看作是这样的一个案例。他们的研究数据来自于“全美第8年级教育跟踪研究”(NELS-88; Ingles, et al., 1992)。在这一全国性的样本中,研究者发现有一些拉丁裔8年级学生在8—10年级之间辍学。关于这一群体内的每个成员,都可以获得各种关于其态度、成绩、家庭背景以及其他方面的资料。在这些资料的基础上,这几位研究者分别对男性和女性的拉丁裔青年建立了形描。这些形描的基础是这两组人每组的最常见特征、平均分(如自我认知及自我效能感分数),以及对资料和文献的总体推论。

比较式形描。这种类型的叙述性形描是对比一个分析单位与另一个分析单位,进而分析二者之间的异同而得到的。例如,Fals-Stewart, Birchler, Schafer 和 Lucente(1994)进行了一项研究,他们从102对寻求救治婚姻的夫妻中,发现了5种不同的夫妻类型。这些类型的差异是基于根据一项临床/个性测试——“明尼苏达多阶段个性目录(MMPI)”的得分比较而形成的。研究者进行了定量分析(整群分析),确认那些彼此相似且与其他夫妻组群不同的夫妻。由此获得的5个夫妻组群被标名为冲突型夫妻、消沉型夫妻、不满型夫妻、烦躁型夫妻以及家庭和睦型夫妻。在另一个复制的样本研究中,研究者也发现了这5种夫妻类型。

通过对这些夫妻组群的辨识,研究者对每个组群的夫妻构建了一个叙述性形象。例如,对“消沉型夫妻”是这样描述的:

> (这种类型)通常是在测试时总体上显得焦虑、担忧和消极的配偶。这种类型的配偶表现出兴趣狭窄、挫折承受性差、情绪低落,并且常常缺乏自信。较之于其他组群(第5种组群除外),这种类型的夫妻并没有太多的不一致之处,也没有对两人间的冲突陷入过度反应。但是,他们对于婚姻的不满程度相对较高。(Fale-Stewart, et al., 1994:235)

规范式形描。这种类型的形描与比较式形描相类似,但其中所进行的比较立基于某一标准。这种“标准”可以是标准化的样本,也可以是一个特定的总体。这种类型的形描在临床和个性心理学、精神病学以及其他相近的学科中得到了非常普遍的应用。通过研究

对象在诊断测试或个性测试(如 MMPI)中所得的分数,研究者可以建构出总体性的印象和叙述性描绘。这些印象式描述的基础是将个体所得的分数与规范性组群的分数所进行的比较研究。

规范式形描研究的案例可参见 Wetzler,Marlowe 和 Sanderson(1994)的研究。在这项研究中,他们对消极的病人进行了一系列临床/个性问卷测试。每名病人都根据他们的测试得分与测试规范分数的对比(MMPI vs. Millon)而确定了相应的"代码类型"。随后,研究者对每一种类型又建立了叙述性形描。例如,MMPI 上的"无能为力的消极"(incapacitated depressive)类型是这样表述的:

> (他们)迷惑、低效,无法找到解决问题的办法。他们的记忆力和注意力也都有缺损。他们对遭遇拒绝十分敏感,与他人保持着情感距离,但同时又与他们自己的感觉和反应相互背离和断裂。结果,他们很难配合治疗。心理疗法应当以标示出他们的感受为目的——用以确定不同状态下的人际需求,并建议他们如何做出最好的反应。(Wetzler,Marlowe and Sanderson,1994:762)

尽管形描清晰易懂,可以提供特定分析单位的"典型"形象,但仍应当慎用。因为即便形描建立在较多的变量之上,仍可能过于简化了群体的形象,其次,绝大多数形描都假设组群具有同质性,因此会着力收集情态式、平均式的回答或信息;最后,在形描的构建过程中,调查者的主观分析可能会影响到研究结论。整体式形描在这一点上尤其突出,在这种类型的形描中,调查者的印象构成了描述的基调。倘若形描是从总体上对许多具体定量信息所进行的定性化总结,并且在运用过程中结合了其他的定性和定量信息,那么,出现这种偏见的风险就可以减少。正如本书第 2 章所提到的那样,建构主义者认为这种偏见是在对定性资料(或其他任何类型的资料)进行阐释的过程中一个不可避免的特点。

顺序型定性—定量分析

在顺序型定性—定量分析的资料分析策略中,首先将对定性资料进行分析,从而确定在某一方面相似的个体组群;随后,再根据已有的定量资料对已确认的组群进行比较,或者根据定性分析之后另

行收集的资料进行比较。这种顺序型策略包括下述一些变体：

第一种是建立人或环境的组群，这是依据定性资料/定性观察而建立的，随后根据定量次数再对这些组群进行比较。按照 Caracelli 和 Greenede 的说法（见框 6.1），我们可以将之称为“类型学建构”。在“类型学建构”的过程中，研究者首先将个体划分到不同的类型之中，随后，根据已有的或另行获得的定量（或定量化）的数据对这些类型组群进行统计学意义上的比较研究。

例如，研究者根据在教室内的观察中所做的现场笔记，可以将教师分类为高绩效组和无绩效组。随后，研究者可以根据定量的变量/测量对这两组教师进行比较分析，如根据教师对问卷工具的回答，或是他们的学生所获得的考试成绩，等等。比较分析的操作可以包括对变异或共变的单变量分析（或多变量分析）、区别函数分析（discriminant function analysis）或其他的统计工具。例如，区别函数分析的研究结果是识别出将不同的组群（如高绩效的教师与无绩效的教师）区分开来的各个变量，并识别出可以将不同的组群区分开的最佳统计指标。

高级非参数工具也可以用来比较分析不同的组群。例如，在一个有两个组群的研究案例中，研究者可以将组群成员作为因变量或预测变量进行逻辑回归分析（其中，预测变量可以由类型变量和连续性变量组成）。分析结果是识别出将两个组群区别开来（即预测组群的成员）的最佳变量。对于超过两个以上的组群，也可以使用这种比较方法。

第二种是先通过内容分析建立不同组群的属性/主题，然后通过另行收集的（或已经获得的）定量数据进行统计分析来加以确证。例如，首先用常数对比分析从定性资料中建立起生成性的主题；然后建立起与之“契合”（且与其他类型差异明显）的各个主题、变量或情势的类型范畴（即概念识别（construct identification））；接下来，研究者用（后来另行收集的或已经获得的）定量数据进行统计分析，对此前定性分析中所获得的推论进行确证或扩展（即概念确认（construct validation））。

再如，研究者可以将教师的叙述（从焦点小组座谈中获得）进行分类，得出体现“好校长”的不同方面的各个主题。这些生成性主题或范畴是组成总概念“校长绩效”的亚概念指标。这些范畴

类型是基于教师的认知和观念中的相同点(或不同点)而形成的。然后,研究者设计出将这些主题群包含在内的问卷调查工具,并以此对一组教师实施问卷调查。接下来,研究者对所获得的定量数据进行要素分析,以确定此前各个定性范畴的效度之等级。

Iwanicki 和 Tashakkori(1994)的研究也是这种分析策略的一个研究范例。在这项研究中,他们首先通过对定性资料的内容分析得到了“高效校长的熟练性”的不同类型;然后,他们向学校校长们发放问卷调查工具,再次对这些类型进行测量;随后,将由此所获得的数据用来进行确证性的要素分析。在相应的定量研究中,大多数先前已区分的“熟练性”类型得到了确证,但有两种“熟练性”从一种类型改换到了另一种类型。同时,他们还根据定量分析进一步提炼出关于“熟练性”的亚类型。

顺序型定量—定性分析

与前面所讨论的顺序型定性—定量分析相似,顺序型定量—定性的分析策略是先在定量数据的基础上确定人和环境的不同组群,然后通过(后来收集的或已获得的)定性资料对这些组群进行比较研究。例如,可以用这种分析策略来建构类型学:先利用从测量教师的认知效能和控制点的调查所获得的数据进行定量分析;然后依据对教师的问卷调查回答的定量分析,将教师划分成 4 个组群(由“高效能—低效能”与“学生的成功源于内部—源于外部”交叉而形成的 4 个区间分类);随后,研究者对这 4 组教师进行观察,并根据由此所获得的定性资料进行比较研究(见 Taylor and Tashakkori,1997)。

最常见的先定量后定性分析的研究案例,是先根据多元回归分析中的剩余值或协方差分析(ANCOVA)中的修正协变值而确认出个体/单位,然后对之进行定性分析;接着,研究者对这些个体/单位在定量分析中获得高分值(或低分值)的原因进行调查,收集到详细的相关资料;然后,研究者对定性资料进行内容分析,或将定性资料转化为定量数据进行统计分析。

还有一个体现这种先定量后定性分析策略的研究案例:研究者先采用回归剩余值的标准化测试,依据测试结果将学校划分成“高效的”和“无效的”两种类型(见 Kochan, Tashakkori and

Teddlie,1996);随后,他们对这两类学校进行观察和相互比较,以探索二者在其他方面(如学校氛围方面)可能存在的差异。

通过定量分析对属性/主题进行分类,然后再通过对其他资料的定性分析来证实这些类型,这种分析策略与前面所提到的概念识别和概念确认的分析过程相类似。在这种策略中,研究者的目的是先通过对定量数据的要因分析,识别出一个概念(子概念)的构成成分,然后收集定性资料来确认所划分的类型,或在有关这一子概念的已有信息的基础上进行扩展。例如,这种混合资料分析的一个研究案例即涉及这样的程序:研究者首先对学校工作人员的一个样本进行问卷调查,并对从中获得的调查数据进行要素分析,进而对教师关于学校氛围之理解的若干维度作出分类。随后,研究者运用观察或其他类型的资料(如焦点小组访谈)来确证这些维度,或者探寻这些维度在日常互动中所呈现的程度高低。

Caracelli 和 Greenede(1993)曾对这种分析类型的又一种运用方式进行过讨论。与上述案例不同,他们的运用这种策略的目的不是对概念确认的结果进行证实或扩展。相反,他们的目的是为接下来的定性/范畴分析提出一个初步的分析框架。例如,可以将要素分析的结论当作本章前文所界定的常数对比分析的起点,即通过要素分析所获得的事件/观察的范畴,可以随后用于在常数对比分析中对此前的定性资料进行编码。

混合模型研究的运用、案例和未来的方向

第三部分

在本书的这一部分，我们将详尽说明若干当下正在进行的和已公开发表的采用混合研究路径的研究案例。我们将在第 7 章概要介绍两个选自学术文献的案例（Ulin, et al., 1996; Patton, 1990），然后我们将首先对确证性研究的各种类型进行讨论——正是在这类确证性研究中，某些类型的实验性概念的预料或期待得以生成。接下来，我们将介绍若干有关探索性研究的案例。这类研究不存在任何事先的预设。正如 Dooley（1995:264）所指出的，在这种研究类型中，“研究资料导出假说，然后我们可以用通常的确证方法对这些事后假说进行检测”。

第 8 章将描述 3 个完全属于混合方法的研究项目，包括一个使用了表 3.1 中所展列的所有 8 种纯粹的和混合的设计方法进行跟踪性研究的完整案例。我们详尽说明这些案例的目的，是要展示通过复杂的混合模型设计所能收集到的信息的广度。

第 9 章则总结了我们在考察社会和行为科学研究中的混合方法和混合模型后所获得的结论。同时，本章还进一步讨论了与这些混合方法和混合模型相关的未来发展问题。

7

混合模型设计的案例

正如我们在本书第 3 章所讨论的，从混合模型类型 Ⅰ 到混合模型类型Ⅳ，都不同于 Patton（1990）所提出的原型设计（prototype designs）。Patton 将研究设计划分成实验型研究设计与自然型研究设计两类，与此相反，我们对研究设计的类型划分是考察它们在多大程度上是确证性研究（包括定性案例研究、实验设计和非实验设计），还是探索性研究（包括自然型研究以及定量探索性研究，如问卷调查）。在本章和下一章中介绍了若干有关这些研究设计的案例。

应当注意的是，框 7.1 所展列的 Patton 的案例是我们所指称的混合模型设计类型 Ⅰ 、类型 Ⅱ 、类型 Ⅲ 和类型Ⅳ的具体应用。以下是这几种混合模型设计类型与框 7.1 所载类型的对照：

混合模型设计类型 Ⅰ：确证性研究、定性资料、统计分析（框 7.1 中的第 2 种形式）

混合模型设计类型 Ⅱ：确证性研究、定性资料、定性分析（框 7.1 中的第 1 种形式）

混合模型设计类型 Ⅲ：探索性研究、定量资料、统计分析（框 7.1 中的第 4 种形式）

混合模型设计类型Ⅳ：探索性研究、定性资料、统计分析（框 7.1 中的第 3 种形式）

Patton 的分类没有将研究设计类型Ⅴ和类型Ⅵ的案例包括进来，因为这两种类型十分少见，且涉及定性化工具（见本书第 6 章）。

框 7.1 Patton 所提供的"混合形式"的评估研究案例

对一个为有可能卷入刑法审判系统的高风险学生提供服务的项目进行评估时,有下列几种可能的方案:

(1)混合形式:实验型设计、定性资料和内容分析。与纯实验形式的研究一样,研究者会将潜在的参与者随机分配到实验组和控制组之中。研究者在研究开始时,先对实验组和控制组中的所有青年进行深度访谈,访谈的关注点与纯定性方法中的关注点是一样的。在研究项目的最后,研究者再次进行访谈。然后,研究者对来自控制组以及实验组的资料分别进行内容分析,并随之对在控制组和实验组中所发现的不同模式进行比较和对照。

(2)混合形式:实验型设计、定性资料和统计分析。研究者将参与者随机分配到控制组或实验组之中,并在项目展开之前以及项目的最后对参与者进行深度访谈。这些访谈资料的原始形式随后被交给一个裁判小组,由他们根据几个结果维度对每次访谈进行评估,在一个 10 进位制的量表上打分。

无论是"事前"访问,还是"事后"访问,裁判的评分涉及这样一些维度:在学校成功的可能性(低 = 1,高 = 10)、刑事犯罪的可能性(低 = 1,高 = 10)、教育的信念、从事生产性工作的信念、自尊、对理想的营养和健康习惯的表达。随后,研究者使用推断统计对这两组人员进行比较。裁判在不知晓参与者属于哪一组的情况下进行评分。评分量表所得的结果还用于统计分析,探讨与参与者背景特征的关联性。

(3)混合形式:自然型研究、定性资料和统计分析。与纯定性形式的研究一样,研究者依据任意一个选用标准,抽取出学生参与研究项目。研究者在研究展开之前以及研究的最后,对每个学生进行深度访谈。随后,这些资料交给一个裁判小组,他们在与前面案例中所列维度相似的维度之上进行评估打分。研究者计算每一位学生的变化分值,并对这些变化进行统计分析,与学生的背景特征关联起来,以便用回归分析的形式确定出具有哪种特征的学生可能会在本项目中预测成功。此外,研究者还会根据用以将活动属性的氛围定量化的一组量表,对项目中观察到的活动进行评分:诸如项目活动所引起的积极参与或消极参与的程度,师生互动的高低程度,互动形式的正式或非正式程度,以及参与者对项目活动的投入程度,

等等。研究者在定性描述的基础上对项目活动进行评分,然后将评分加总起来产生出对项目实验环境的总体概观。

(4)混合形式:自然型研究、定量资料和统计分析。研究者根据自己的标准抽取出参与项目的学生。评估者在进入项目场景之前,不带有任何预先确定的分析范畴,或者有关重要变量和变量间关系的预先假定。评估者对项目中的重要活动和事件进行观察,从中寻找出行为和互动的类型。对每一种行为或互动的新形式,评估者都创建一个类型,然后使用时空抽样设计来计算这些行为、互动类型出现的频率。随后,研究者将这些观察到的行为和互动的出现频率,与群体规模、活动持续时间、工作人员—学生的比例以及社会密度/物理密度等特征建立起统计关联。

来源:Patton(1990:191-193)。

注意:以上是评估有可能卷入刑法审判系统的高危险学生的研究项目时,存在的几种可能的解决方案。

确证性研究

在表3.1中,我们将混合模型的确证性研究设计归类为混合模型设计类型Ⅰ、类型Ⅱ和类型Ⅴ。就传统而论,在确证性研究中,研究资料主要是定量的,概念框架是演绎的,而资料分析则是统计学的。但是,在我们这里所讨论的混合确证性设计中,研究所采用的资料既可以是定性的,也可以是定量的,研究者也可以使用两者中的任一种形式对资料进行分析。在使用这类研究设计的研究课题中,至少也会作出尝试性的预估并在研究中进行检验。下文就是对这一组别的三种类型设计的讨论。

类型Ⅰ:确证性研究、定性资料和操作、统计分析和推论

在这种类型的研究中,所收集的资料是定性的,研究者对这些定性资料进行定量化转换,然后纳入统计分析。这也是学术文章中最常见的混合设计之一。Pemberton、Insko和Schopler(1996)关于集体竞争力与个体竞争力的对比研究,就是这种研究设计的典型案

例。在这项对比研究中，研究者首先进行了一次实验（详见 Pemberton，et al.，1996：954）。实验表明，大学生对集体间竞争性互动的记忆，要多于对个体间竞争性互动的记忆。在这一实验之后，几位作者对集体间互动会被认为更具竞争性的原因提出了疑问。他们在相关理论和前人研究的基础上提出了试探性的解答：①参与者意识到实验的预期（在本书第 4 章和第 5 章的参与者反应和参与者角色的章节中对此作过讨论）；②存在着一种外围集体的心理图式；③记忆效应（以往的集体经历与体育竞赛有关，而其中的竞争是一种常态）。

在接下来的类型 I 的研究中，Pemberton 和他的同事们使用了“罗切斯特互动记录（Rochester Interaction Record，RIR）”（Reis and Wheeler，1991）的资料收集程序来记录自然环境中的互动。研究者试图阐明和检验许多假说和预测。例如，研究者预测，研究的参与者会回忆起更多的集体竞争事件。另外，研究者还提出预测，对 RIR 记录的真实事件的观察，将会显示出集体间互动比个体间互动具有更加激烈的竞争性。

实验的参与者是刚入门的心理学学生，有 28 位男性和 27 位女性。作为 RIR 程序的一部分，研究者向每名参与者发放了一本 40 页的小册子，每一页记录一次互动。要求每位参与者记录所发生过的互动的时间，并画圈标出互动的类型（如个体对个体、个体对集体、集体对个体、集体对个体等）。

这项实验的一个有趣特征是，参与者本人就是将定性资料（各次互动）定量化的评估者。在记录的时候（即每当参与者记录事件时），参与者需在一个有 7 个等级的量表上对每次所记录的事件进行评分，而评分等级是关于互动的竞争性或合作性的程度。随后，研究者对这些定量化的资料进行统计分析和报告。

Smith、Selles 和 Clevenger（1994）所作的研究是这种设计类型的又一个典型案例。研究者们对婚姻咨询过程中的夫妻习惯展开了调查，他们收集了定性资料，将其中一部分定量化处理，然后使用卡方检验和 phi 相关系数对这些资料进行了定量分析。框 7.2说明了这一研究的简要内容。

框7.2 混合模型研究类型Ⅰ的案例

Smith 等人(1994)的研究是一项确证性研究,涉及对参与了婚姻治疗疗程的夫妇进行非结构性的访谈。在此前的研究中,他们已经收集了包括对治疗中的夫妻以及其治疗师进行访谈的定性资料。对这些资料的定性分析产生了两个主题:第一,对于夫妻习惯的反思,治疗师与夫妻的认知是有差异的;第二,这种差异与夫妻和"客户指数的治疗师系统"之间的"空间距离或边界"相关(Smith, et al. ,1994:269)。新启动的这项研究的目的是对基于这两个判断之上的假说进行确证研究。

研究者用了4个月的时间对夫妻和治疗师进行民族志访谈,收集定性资料,而先前研究中所得到的区分类型则被用来对这些资料进行初步的定性分析。内容分析产生出了经过完善的7种类型。研究者对每个有关主题/认知的这些类型都记录了发生的频率,并进行了定量分析。

对于第一个假设,研究者使用了卡方检验来比较夫妻认知与治疗师认知在每一种类型中的频率。结果显示,这两组人谈论性别问题的频率是一样的。但治疗师最关心的是疗程中夫妻对习惯进行反思的效果,而夫妻最关心的是这些习惯的好处。对于第二种假设,研究者们通过计算"空间距离"与"倾听过程"之间共变的 Phi 相关性来进行测量。研究者们发现二者之间存在着强相关关系,"在整个文本中,有许多案例证明了空间距离与倾听过程之间存在着共变关系"(Smith, et al. ,1994:280)。

类型Ⅱ:确证性研究、定性资料和操作、定性分析和推论

这种混合设计基本上是一种确证性的定性研究。与传统的自然型设计相反,这种研究设计至少在某些方面采用了尝试性的预测。这种类型中最常见的研究课题就是使用定性资料的收集和分析来确证或否证此前的研究发现。

Sinclair(1994)的研究是以这种设计类型为主要模式的一个范例。Sinclair 使用定性资料和程序,就科学课程中以进行预测作为教学法的效果提出了4个假说并进行检验。参与研究的学生被分成两组:一组是在遗传学教学单元中加入了预测活动的实验组,另一组是在遗传学教学单元中使用传统方法的控制组。研究

者收集了学生访谈、教师访谈、教师日记和教室观察作为定性资料。在对研究资料所作的定性分析中，研究者采用了 Bogdan 和 Biklen(1982)的常数对比法，将各种反响分成不同的类别。

定性分析显示，在教学中进行预测的活动使学生们的参与程度和兴趣程度都有较大的提升。实验组的学生在教学中参与更多，提出了更多"深思熟虑"的问题，并且对课程的兴致看上去比那些控制组的学生更高。观察者还报告说，这些实验组的学生在教学中与教师之间你问我答的对话，也要比控制组的学生更多。定性分析还表明，教师的教学风格的差异影响到两组学生所在课堂的进行方式，并且可能影响到学习的结果。

关于这种设计类型的另一个案例是 Mann(1994)的研究。他收集了关于两组辅导员的定性资料。实验组是 3 名看上去已经从培训中受益最多的辅导员(他们在成果测量中得分最高)，控制组则包括了那些在培训项目中没有受益的辅导员(成果测量上的分数很低)。总的假设是，对于他们自己的角色以及对他们加之于受教者之影响的理解，这两组辅导员是不同的。

研究者所收集的定性资料包含在 10 周时间内每周的流水记录，以及每位辅导员所写的两页纸的短文。对这些流水记录的定性分析显示，得分高的辅导员比得分低的辅导员对教学过程中的人际交流更有意识，对冲突的辨认更为清晰，在解决冲突上也更有效率。得分低的辅导员对整个实验期的评估忧心忡忡，对受教者的问题作出了太笼统的概括，相对于高分数的辅导员，他们更为焦虑，也更难成功。(关于这些流水记录的更详细案例，见 Mann，1994)

类型Ⅴ：确证性研究、定量资料/操作、定性分析和推论

尽管这种类型的设计并不常见，但在若干领域的研究文献中仍能见到相关的研究案例：定量资料被定性化，并以定性的方式进行表述和分析。

类型Ⅴ的典型性研究课题是在定量资料的基础上，对群体、个体的"形象"进行辨识和文字表述。这类研究设计的一个典型案例是，有研究者设计了一项研究，旨在以访谈(问卷调查)资料为基础，对扎伊尔某个城市中心使用避孕工具的妇女的不同组群

或“类型”之属性进行确认。研究者在此前研究的基础上提出了一个类型学(如持续使用者、偶尔使用者),也根据调查资料为每一种类型或组群描绘了一个整体的形象(见本书第6章)。关于这种设计类型的案例见本书第8章。

探索性研究

混合模型的探索性研究设计在表3.1中被标记为混合模型设计类型Ⅲ、类型Ⅳ和类型Ⅵ。这些混合模型的探索性研究使用定性资料或定量资料,并使用定性与定量中的一种路径对这些资料进行分析。下面我们将对这一组的三种不同的设计类型进行分析。

类型Ⅲ:探索性研究、定量资料和操作、统计分析和推论

这种类型的研究设计与传统定量的探索性研究(通常称为描述型研究)相同,即在研究之前先不作任何预测。将这种研究纳入混合模型研究的范畴,是因为它们并不符合我们在表3.1中展列其逻辑框架的假说—演绎式的传统研究。传统的描述型研究设计与混合类型Ⅲ设计的平行相同典型地体现了我们将在本书第9章所展开分析的一种现象:在传统领域(如心理学、人类学、社会学)中治学的研究者可能会使用混合模型设计,但由于他们在定量或定性传统之中浸润太久、太深,以至于他们会自认为只是在进行“纯粹”的定量或定性研究。

Aghajanian和Moghadas最近所作的一项探索性研究属于这种类型Ⅲ设计的典型案例。研究者对伊朗近年来的离婚趋势进行了研究,试图探讨其中潜在的决定因素以及各种离婚样式的后果。与其他探索性研究相同,他们没有进行任何预先假设。研究资料是通过在一个大的城市中心进行问卷调查而收集起来的。研究者在1986年人口普查的基础上,对已婚妇女和离婚妇女进行了多层次抽样。在研究的第一阶段,研究者从169个普查区随机抽取出43个普查区,然后从普查资料中辨认出所有至少有一名离婚妇女的家庭。结果就得到了一个由254名妇女组成的样本,她们都曾经结过婚,后来又离婚了。研究者也抽取(2%是随机抽取

的）了一个比较性的样本，这一样本由那些如今已婚并且没有离婚的妇女组成。女性访谈者拜访了每一户家庭，并对已确认的妇女进行了访谈，收集了有关社会经济特征、态度、健康和婚姻问题等多方面的详细资料。

研究结果显示，由于社会法律的变迁以及持续几年的对伊拉克战争的影响，伊朗的离婚率发生了一些变化。研究者发现，两组妇女（离婚妇女和未离婚妇女）间最主要的区别集中于城市化特性、教育背景、工作地位和宗教信仰等方面。离婚妇女中更大的一部分来自城市中心，学历或高或低（与中产阶层相比），宗教忠诚程度也更低。在离婚结果方面，研究者发现，离婚妇女承受着经济上的痛苦，比没有离婚的妇女经历着更多的心理问题。而且，离婚妇女的孩子比没有离婚妇女的孩子经历了更高程度的情感问题，少年犯罪率也更高。研究者的结论是，在伊朗，离婚的结果要比在美国严重得多。此外，在美国发现的首婚年龄与离婚可能性之间所存在的反比关系，并没有在伊朗发现。

类型Ⅳ：探索性研究、定性资料和操作、统计分析和推论

在心理学和社会学研究中，有大量的研究使用投射技术（projective techniques）（完型填空、故事叙述等）来收集资料。在完成资料收集工作之后，研究者会通过不同的程序将资料定量化（见本书第 6 章），并借助若干工具（大部分是非参数的）如线性工具对数模型分析（log-linear modeling）和逻辑回归等展开分析。

类型Ⅵ：探索性研究、定量资料和操作、定性分析和推论

这种类型的研究设计也较为少见。在这种研究设计中，研究者会收集定量资料，然后对之进行定性分析和表述。本书第 6 章在讨论定量资料定性化时，曾列举了有关这种设计类型的几个案例。在这种类型的研究中，研究者往往会根据定量信息而将参与者分成不同的组群（如极端案例组、参与者的类型）；随后，研究者再采用定性（叙述）形式对这些组群进行描述（如我们在第 6 章所讨论的“形象描绘”）。

Rusbult 等人（1993）所作的研究（本书第 8 章将有详细的讨论）是类型Ⅵ混合模型研究的一个典型案例。在这项探索性的研

究中,要求参与者对一系列关于罗曼蒂克关系的描述进行评分,评分的依据是这些描述与一系列"目标"描述之间的相似程度。根据多维度量表分析(多变量统计分析的一种),研究者分辨出了这些描述中的4种关系类型。然后,研究者在定性和定量信息的基础上对这4种原型进行详细的描述。框7.3列举了两个关于这种原型(形象)的案例。

框7.3 Rusbult 等人研究中的不同原型(形象)的案例

如画似梦(女性):象限1中的心理模型,在亲密程度上得分不高,但在浪漫—传统主义上得分很高。该类型是高度理想化的,形成过程快捷而自然,在投入上有明确的要求,不是平等主义的(通常扮演传统的性别角色),并不看重友情,不是建立在尊重/景仰的基础之上,并且往往是男性采取主动。这种关系中的主体并不是寻求信任和亲密关系,而是渴望经典的浪漫幻想——迷人的举止,精彩的生活和澎湃的感情。在这里,一见钟情是常见的主题。我们将这些理想称为"如画似梦",下述摘录展现了这些理想的特征:一见钟情;完美爱人;在性的方面,他从不强迫我,他在这一方面很成熟,与众不同……我爱得神魂颠倒,感到前所未有的幸福。

伴侣(男性):象限4的心理模型,是象限1的心理模型的对立镜像。它的特征是:日久生情,平等主义,朋友取向,务实共处,相敬如宾,男方并不主动,对二人关系缺乏明确的要求。这是一种非常舒适的、低调的理想——恋人是密友和同盟,享有共同的事业兴趣、政治信念和生活关注。下述摘录描述了象限4的伴侣关系:以信任和回报为中心;彼此认真对待,我不喜欢轻佻游戏;诚实非常重要;而且,我喜欢有进取心的女孩,我喜欢承担所有的任务和危险……她可以与其他男人说话,我不会为此嫉妒或反应过度;她爱我是因为我这个人,不是因为我有什么、我与谁有关系或我的交往圈子。

来源:Rusbult 等人(1993:507-509)。

这种设计的另一个案例是我们将在第8章讨论的 Taylor 和 Tashakkori(1997)的研究。这两位作者在一个重建的学校地区中抽取了一个教师样本,从中收集问题调查资料。研究者没有作出先验的假设。他们的调查包含了两种特定类型的问题:关于教师参与决策的愿望和动力,以及他们对自己在教学和学校事务上实

际介入决策的自述报告。研究者根据教师在两个维度(参与动机与实际参与程度)上的得分值,建立了四个极端教师组。由此,研究者在这个双维度模型中分辨出了4组不同“类型”的教师。这4种类型分别被命名为“有权型”(想参与并确实参与了)、“无权型”(想参与却并未参与)、“卷入型”(不愿意参与却进入了决策过程)和“无关型”(不愿意参与,也并未卷入)。根据其他一些有关各组教师的信息,研究者进一步建构了文字性(定性)的典型形象。

8

混合模型设计的扩展案例

正如我们在前面所讨论的，完全的混合研究同时采用两种类型的资料收集技术(定性和定量)和两种类型的资料分析路径(统计和定性分析)，其中一些完全的混合研究还可能将探索性方法和确证性研究结合起来。框 8.1 通过图表向我们展示了在一个对妇女生育健康状况的国际研究中所使用的混合模型方法(Ulin, et al. ,1996)。如图所示，混合的程度和类型可能会在这一大规模研究的不同子项目中有所不同。

我们将在本章提供另外 3 个关于完全混合研究的扩展案例，最终目的是为研究生和感兴趣的研究者提供可用作模板的完全混合型研究的更多具体案例。

在表 1.1 中，我们将这些研究类型称为在各个研究阶段(研究的类型，或资料收集/操作，或分析/推论)上多元使用的混合模型研究。此外，表 1.1 的注释进一步说明了方法的混合使用必须使定性和定量两种方法至少在研究的一个步骤中同时出现。本章所提供的案例在资料收集/操作阶段以及分析/推论步骤上都使用了混合方法。

第一个案例(Taylor and Tashakkori,1997;Taylor,Tashakkori and Hardwick,1996)是一个平行的混合模型研究，这项研究以平行的形式在资料收集、资料分析和推论过程中将定性和定量两种方法结合起来(多元同步方法)。推论过程同时在演绎和归纳逻辑上展开，而资料的收集和分析则同时使用了定性路径和定量路径。

第二个案例(Rusbult,et al. ,1993)则有所不同，其研究的每一个步骤都依赖于前一步骤所产生的发现、概念发展或素材。因此，

框 8.1　一项多国研究中的混合模型/设计

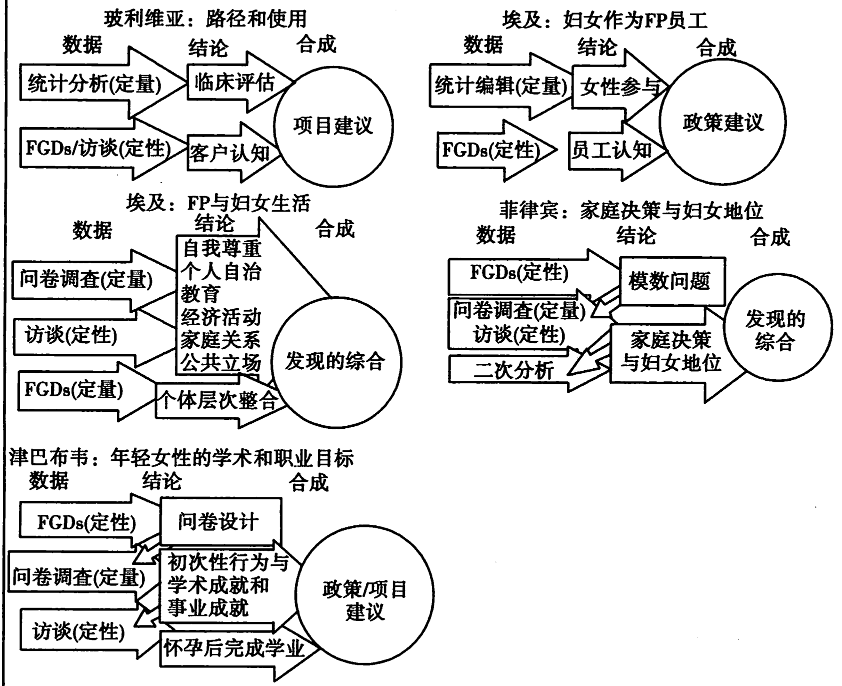


表 8.1　复合混合模型设计：类型Ⅶ和类型Ⅷ

类型Ⅶ：平行混合模型研究*	类型Ⅷ：顺序型混合模型研究**
单阶段研究	研究的第一阶段
第一步：研究的类型：定性和定量	第一步：研究的类型：定性或定量
第二步：资料收集和操作：定性和定量	第二步：资料收集和操作：定性或定量
第三步：分析和推论：定性和定量	第三步：分析和推论：定性或定量研究的第二阶段：
	第一步：研究的类型：定性或定量
	第二步：资料收集和操作：定性或定量
	第三步：分析和推论：定性或定量其他的重复阶段

＊方法的混合是指至少在研究的某一步骤同时存在两种路径。

＊＊方法的混合是指每一种方法至少在研究的某一阶段出现。

这项研究理当视为顺序型混合模型研究的一种典型。在这类研究中,研究者分步骤地依次使用多种资料收集、分析和推论的方法。每一步骤本身可能会使用一种混合方法,并为相衔接的下一研究步骤提供概念和(或)方法论基础。

第三个扩展案例是一个长期项目——路易斯安那学校效率研究(LSES;Stringfield and Teddlie,1990,1991;Stringfield and Teddlie,1993;Teddlie,et al.,1989)。这项研究分为多个阶段,每一阶段都结合使用了多种混合模型设计。我们将这一项目作为混合模型研究的最终的、详细的分步示例,其中,表3.1中所举的8种研究类型(纯粹模型设计和混合模型设计),以及这些设计的更复杂的结合形式都得到了应用。

表8.1中对平行型混合模型研究(混合模型类型Ⅶ)和顺序型混合模型研究(混合模型类型Ⅷ)作了极其详尽的介绍,而表8.2则对一项研究的步骤与阶段作了区分,并列举了包括有这些区分的一个研究案例。正如本书第1章所界定的,步骤(stage)是指研究的一个组成部分。例如,在大多数研究中,研究步骤包括形成研究问题、确定研究类型、资料收集、资料分析和最终推论。阶段(phase)则是指由许多步骤构成的一个完整的研究过程,每个阶段都是整体研究的一个部分。例如,一项研究可能有一个以定性为主的阶段(包括设计、资料收集、资料分析和推论等步骤),然后接着一个有着相同步骤的、以定量为主的阶段。

平行型混合模型研究(类型Ⅶ)在研究的每一阶段上都有方法的混合。在这类研究中,研究者至少在研究的某一步骤(如资料收集)上,会同时使用定性方法和定量方法。顺序型混合模型研究(类型Ⅷ)是指在不同的阶段上有方法的混合;尽管研究的一个或多个步骤中会有某种程度的方法的混合,但在研究的某个阶段中主要使用两种方法(定性或定量)中的一种,而研究的下个阶段则会使用另一种方法。例如,如果第一阶段占主导地位的是定性方法,那么,第二阶段占主导地位的就是定量方法。而且,每个阶段的设计都包括对前一阶段中所提出的问题进行探索或确证。

表 8.2　路易斯安那学校效率研究(LSES)的五个阶段

阶　段	简　介	时　期
LSES-Ⅰ 前导研究	●1980—1981 年项目的概念化 ●1981—1982 年的前导研究,包括对工具进行田野测试 ●完成 LSES-Ⅰ报告	1980— 1981
LSES-Ⅱ 宏观层面的研究(过程—产品研究)	●抽取 76 个学校的样本 ●1982—1983 年从 74 个校长、250 位教师及超过 5 400 学生中收集关于学校氛围的问卷和其他工具 ●分析资料 ●1984 年完成 LSES-Ⅱ报告	1982— 1984
LSES-Ⅲ 微观层面历时性研究(案例研究、第一轮现场)	●抽取 8 个配对学校组的样本 ●1984—1985 年进行了超过 700 小时的课堂观察和超过 1 000 小时的现场资料收集 ●分析数据 ●使用 LSES-Ⅲ定性资料和定量数据准备好研究报告和文章	1984— 1988
LSES-Ⅳ 继续进行微观层面历时性研究(案例研究、第二轮现场走访)	●1988—1989 年设计 LSES-Ⅳ ●1988—1990 年进行了超过 650 小时的课堂观察和总体超过 950 小时的现场资料收集 ●分析数据 ●使用 LSES-Ⅲ和 LSES-Ⅳ数据准备研究报告和文章	1988— 1992
LSES-Ⅴ 继续进行微观层面纵向研究(案例研究、第三轮现场走访)	●1993 年发表的《学校造就差异:来自一项对学校效率的十年研究的经验教训》一书,总结了从 LSES-Ⅰ到 LSES-Ⅴ的研究过程 ●1994—1995 年设计 LSES-Ⅴ ●1995—1996 年进行了超过 500 小时的课堂观察和超过 750 小时的现场资料收集 ●开始展开资料分析	1992— 1996

一个平行型混合模型研究(类型Ⅶ)的案例

Taylor 等人(1996;Taylor and Tashakkori,1997)曾经作过一项关于学校重建效果的研究。在这项研究中,研究者以一种完全混合研究的形式将定性和定量的资料收集、数据分析和推论过程结合起来使用。研究者首先要求某一学区内的所有教师回答一项封闭式的问卷调查工具。这一问卷结合了出自 1988 年"全美教育跟踪研究"中的次级量表与出自 Bacharach、Bauer 和 Shedd (1986)所提出来的次级量表,所询问的问题涉及教师对学校氛围、学生纪律、教师效能感以及教师在决策中的参与程度等问题的认知。对这些问题的回答按照李克特量表的四分法模式排成从"非常不同意"到"非常同意"的不同选项。同时,其中还附加了一个开放性部分,要求教师给出他们对学校以及其他相关问题的感觉和观点,以此获得定性的回答。

除了收集问题调查的数据外,经过专门训练的观察者还对 7 所学校逐一进行了课堂观察。在两种不同的场景下,观察者至少对每所学校的 8 个课堂进行了观察。观察者所做的广泛的田野笔记不仅记录了这些课堂观察所得到的资料,而且还包括了每个学校其他部分的各种细节。研究设计的第三个组成部分是焦点小组访谈,这些焦点小组包括小学教师、中学教师、小学学生、中学学生、学校行政人员、学校董事会成员以及学生家长。

研究者进行资料分析时同时采用了定量方法和定性方法。他们对调查资料进行了单变量分析和多变量分析。根据开放性回答、田野笔记以及登录的焦点小组访谈,研究者又进行了内容分析。作为分析工作的组成部分,研究者通过计算相关事件发生的频率而将一些定性资料定量化。同时,研究者依据对问卷工具的不同回答而将教师分成不同的组群,从而将一些定量数据定性化(见本书第 7 章的案例),并分别对这些不同的教师组群建立了形象描绘。最后的研究结果同时以定量指数和定性推论的形式表达出来。

一个顺序型混合模型研究的案例

尽管 Rusbult 等人(1993)并不认为他们的方法论是“混合的”,但他们所作的研究其实就是混合设计的一个有趣案例。之所以有趣,至少有下面三个原因:①作者有效地将实验程序与定性的资料收集和推论结合在一起;②他们展开了一系列顺序型研究,每一阶段都使用了前一阶段的素材和(或)概念框架;③这是能够从诸如第6章所讨论的对定性资料进行精细和系统的、定性的内容分析中得到好处的研究项目之典型。

这项研究立基于对研究文献的初步回顾,包括社会心理学关于喜欢、爱情和浪漫关系的各种理论。在文献回顾之后,研究者(Rusbult, et al., 1993:498)接下来提出了研究的主要问题:“人们对于理想的浪漫关系之预期的想象合乎这些理论吗?”研究者在社会心理学的文献中无法找到有关这一笼统问题的答案,于是他们又提出了4个具体的问题以引导研究进程:①年轻人关于理想关系的心理模型有何独特的维度?(也就是说,关于理想关系的心理模型之间有何差异)②理想关系的主导型心理模型是什么?③所谓的理想关系与年轻人对现实关系或潜在关系的思考方式是类似的吗?④女性和男性对于理想关系的心理模型是否存在着根本的差别?

这一研究按照叙述性研究、定性研究和大体上归纳推论的方向展开。但是,随着新的研究(阶段)的推进,研究者收集到大多属于定量的数据,他们对这些数据进行了统计分析。研究者展开了3项研究以探讨“年轻人有关理想的浪漫关系的心理模型”(Rusbult, et al., 1993:493)。前两项研究都包括了3个阶段。

第1阶段包括定性资料收集、部分的定性分析以及形成用于第2阶段研究的素材。研究者要求参与者(本科大学生)描述出理想的浪漫关系(定性资料),并对由此所收集到的描述进行编辑,使之语句规范清晰,然后将之印在小型的指数卡片上。然后,研究者请多个裁判对每个表述进行评定打分,然后将之加总起来得到关于清晰程度的总体指数。

从那些被判定为最清晰的描述(即总的清晰分数高于中间值

的分数)之中,研究者抽取了50个样本(25个样本为男性的表达,另25个为女性的表达)用于下一阶段的研究(第2阶段保留了35个描述)。其中的10个描述被抽取出来(在报告的余下部分中被称为"靶标")。参与者被要求判断所有其他的描述与这10个"靶标"的相似程度。

研究的第2阶段还包括由回答者将定性资料定量化,并对之进行统计分析。首先要求回答者根据自己与自己所分到"靶标"情景的相似程度,对第1阶段中所获得的定性描述排出等级顺序;还要求他们特别注意描述中所表达的关于浪漫关系的理想模式。在确定描述的等级次序之后,接着要求每一参与者简单地书面概括他/她所使用的区分不同描述的标准(这些定性资料没有得到分析,或者他们的定性分析没有写进报告之中)。研究者对每个可能的描述组别的等级得分之间的差异值进行了计算。随后,这些"不相似"的分值被纳入到多维度的测量中(ALSCAL,见Takane,Young and DeLeeuw,1997)。

基于不同的统计学考虑,研究者确定了一个两维度的模型(结构)以区分不同描述之间的差异性。其中一个维度分成亲密关系对表面关系,另外一个维度则分成浪漫/传统的关系与现实/非传统的关系。根据这种(2×2)结构,又确定出4种理想关系的类型,即如画似梦、幸福婚姻、利益关系和伴侣关系。

体现参与者对浪漫关系之看法的"心理模型"的两个维度,以及在这个两维度模型中所形成的4个象限,并不能根据统计分析而进行标签定名。因此,研究的第3阶段是试图获得关于这些维度和象限的定性阐释和(或)标名。在第2阶段中研究者所形成的39个潜在的标名被用来标注不同的关系结构。这些标名来自于多个来源:参与者在第2阶段的描述,调查者对所获统计维度的阐释,以及社会心理文献中关于浪漫关系的模型和理论。

这项研究(在多个阶段里)确定了浪漫关系的类型以及它们之间可能的差异。Rusbult等人提出的另外一个研究问题是关于由此所确定的心理模型究竟是白日梦的结果,还是等同于年轻人在评估浪漫关系时的实际心理表达。针对这一问题,研究者设计了另一项新的研究。研究的参与者包括25个男性和26个女性,他们都自我认定进入当下的恋爱关系至少已有一个月之久。研

究者要求他们写出短文,对3个详细的指示体系作出回答,这3个指示系列分别是描述他们的:①理想中的浪漫关系;②在浪漫关系中他们寻求什么;③他们目前的恋爱关系。

研究者对裁判进行训练,让他们根据此前的第3阶段研究所获得的类型标名而对每个描述进行评分。研究者再将对于每种类型的描述所得到的评分进行平均值计算,从而获得一个有关这一类型的分值。然后,研究者通过3种情景(即前述指示体系)对这些类型分值进行比较。

差异分析显示,这3个情景之间的差异性非常微小。同时,类型标签也被分入研究者根据此前(第3阶段)的统计分析所得出的4个象限之中。研究者通过相关性分析来评估在这4个象限区间中的类型分值之间的对应程度。基于这些分析,作者们得出了结论:前面的研究所获得的"心理模型"与当前研究所获得的真实(现实)模型是一致的。

在单项研究中分步骤进行简单和复杂的混合模型设计的案例

"路易斯安那学校效率研究"(LSES)构思于1980年,最终的数据收集于1996年春季展开。在这16年左右的历史中,这项研究的进展经历了5个数据收集的阶段,中间穿插着一些分析和撰写报告的工作时期。表8.2对这5个阶段进行了总结,下述的文字将简单描述这项研究的整体设计。

LSES-Ⅰ和LSES-Ⅱ的设计

LSES的第一阶段和第二阶段(LSES-Ⅰ和LSES-Ⅱ)主要是进行确证性研究,而后面的三个阶段(从LSES-Ⅲ到LSES-Ⅴ)则既包括探索性研究的成分,也包括确证性研究的成分。LSES始终是由基本的研究问题所驱动的。但在整个项目过程中,研究者们为了解答这些问题而演变出了一些特定的创新型研究方法。以下关于从LSES-Ⅰ到LSES-Ⅴ的整个设计摘自Teddlie和Stringfield(1993)的著述。

在1981—1982学年,研究者展开了前导研究(LSES-Ⅰ)。在

LSES-Ⅰ阶段上,其完成了两项单独的研究活动:①关于学校氛围的调查问卷,这是通过对 Brookover 等人(1979)所使用的问卷修改而成的。研究者在一个学区里对问卷进行田野测试,然后根据所收集到的信息对问卷进行完善。②LSES-Ⅱ的整套方法在另一个学区里的许多学校进行了前导性测试。在对修改后的有关学校氛围的学生问卷所进行的田野测试中,作者们首先调查了常规课堂环境下的 3 年级学生,随后对分成多个焦点小组的学生进行访谈,以确定他们是否能够理解问卷中的问题,并收集如何改进问卷的信息。

LSES-Ⅱ的研究资料收集工作在 1982—1983 学年展开(Teddlie, et al. ,1984)。研究者为 LSES-Ⅱ所设计的抽样框架意在使研究样本能够代表全国各小学的总体。最后,抽取出的研究样本包括 76 个学校,其中 3 年级的教师超过 250 人,3 年级的学生超过 5 400 人。同时,LSES-Ⅱ覆盖了 12 个学区,分别位于城市、郊区以及农村地区。

LSES-Ⅱ所使用的主要工具是标准参考成绩测试(norm-referenced achievement test, NRT),以及经修改的 Brookover 等人(1979)设计的对学生、教师以及校长的关于学校氛围的调查问卷。学生的调查问卷是在常规安排的课堂上集体完成的,校长和教师的问卷则是由他们本人单独完成的。各研究小组分别进入了抽取出来的 76 所学校,并在一天中完成了资料的收集。研究者还从几个二手来源收集档案资料(如学生社会经济地位的资料、国家标准参考测试的成绩或 CRTs,以及学校的结构特征,等等)。

按照界定,LSES-Ⅱ是一个与 Brookover 等人(1979)在密歇根的研究相类似的大型过程—产品研究(process-product study)。资料分析的重点有两层:①确定由学生的社会经济地位和学校氛围程度所导致的学生成绩的变化程度;②根据学生群体的效率状态和社会经济地位特征的不同状况,对各个学校进行比较。

其中,第一组分析采用了多元回归分析和 HLM 方法(等级线性建模,hierarchical linear modeling)。第二组分析(对变异的多变量分析和单一变量分析,或 MANOVAs/ANOVAs)使研究者可以作出这样一些比较研究:①高效率的学校、一般效率的学校与无效率的学校之间的差异性;②中等社会经济地位的学校和低等社会

经济地位的学校之间的差异性;③由研究设计所产生的 6 个组群学校间的差异性。

从 LSES-Ⅲ 到 LSES-Ⅴ 的设计

研究者设计的这 3 个阶段是用来判断不论学生群体和教职工如何变化,单个学校在长时期内保持其效率的程度。这 3 个阶段提供了在横跨 11 年的时段里对 8 对相互配对的高效率学校和无效率学校所进行的跟踪性案例研究。作为这 3 个阶段的研究对象,这些学校是依据特定的目的而被抽取出来的,用以代表全州不同的地理区域和城市背景。抽样的过程相当复杂,包括 9 个不同的步骤(Teddlie and Stringfield,1993)。资料收集工作在 1984—1985 学年展开(LSES-Ⅲ)。在 LSES-Ⅴ 阶段,研究者分别在 1989—1990 学年和 1995—1996 学年,对同一批学校进行了重访。

研究中使用了大量的工具,其中包括对学生、教师和校长的有关学校氛围的问卷调查工具,针对所有 3 年级学生的"国家标准参考成绩测试",对教师和校长进行访谈的草案,一份有 12 页的学校观察清单,高度推论性的课堂观察工具(用于测量教学的质量),低度推论性的课堂观察工具(用于测量作业时间),由研究者用来对学校层次氛围进行评估的工具,用来评估教员之间关系的社会人际关系测量指数。这些工具包括定性测量和定量测量的混合。

在这 3 个研究阶段,观察者对每个学校都花费了数天时间(在 LSES-Ⅲ 阶段,每个学校 12 个人/日;LSES-Ⅳ 阶段,每个学校 10 个人/日;LSES-Ⅴ 阶段,每个学校 10 个人/日)。从 LSES-Ⅲ 到 LSES-Ⅴ,主要的工作重点是在有效率学校和无效率学校的配对中,收集关于教师的课堂观察资料。因此,对课堂观察的记录,LSES-Ⅲ 超过 700 小时,LSES-Ⅳ 大约 650 小时,LSES-Ⅴ 则为 500 小时。这些课堂观察的结果构成了这一时期研究结果的很大一部分。

在接下来的篇幅中,我们将列举"路易斯安那学校效率研究"(LSES)课题中所涉及的表 3.1 已给出描述的 8 种(即全部)单纯型研究设计和混合型研究设计。

纯定量研究

LSES-Ⅱ沿袭了学校效率研究中所采用的定量取向的教育成果函数研究。主要的研究目的之一,是要确定学生成绩的变化有多大程度可归因于家庭背景(如学生的社会经济地位以及人口统计学方面的其他特征),或归因于学校的影响(如学校的氛围)。这一部分的研究是纯定量的,也是确证性的,因为研究者预先建构了假设,认为学校氛围这一变量能够在很大程度上影响学生成绩的变化,由此证明"学校造就差别";研究资料的收集和操作过程也是纯定量的,采用了纸笔式回答的态度量表、"全美标准参考测试"的数据和数字式档案数据;分析和推论过程也是纯定量的,使用了回归分析或 HLM 统计工具,并根据研究假设和此前的研究成果对结论进行了阐释。

纯定性研究

在跟踪性案例研究(从 LSES-Ⅲ到 LSES-Ⅴ)中,研究者建立了探索性的研究问题和确证性假设,同时使用了定性与定量的数据和程序,也同时使用了定性资料分析和统计分析。因此,就总体而言,这些案例研究就是表 1.1 中我们所谓的"完全混合研究"或"研究阶段内多元应用"的混合模型的案例。

尽管这些案例研究属于完全混合研究,但在这些研究中也存在本质上属于纯定性的独立研究。"路易斯安那学校效率研究(LSES)"课题中纯定性设计的最好案例是由 Stringfield 和 Teddlie(1990)所完成的 LSES-Ⅲ阶段上的"自然发生的学校进步"课题。在这项研究中,研究者"发现"了在学校效率研究中还未曾讨论过的一个现象:一些以前被指称是并确定是无效率的学校处在进步之中,却没有任何外部变化因素与此有关。这一部分最初的资料收集和操作过程都是纯定性的,建立在从田野笔记、访谈、观察和录音带中所得到的叙述性资料的基础上;分析和推论也是纯定性的,研究者单独使用了归纳推理过程,尤其是常数对比方法;与"自然发生的学校进步"过程相关的研究主题也是作为这些定性分析的结果而生成的。

类型Ⅰ 混合研究：
确证性研究、定性资料/操作、统计分析和推论

Teddlie 等人（1989）所作的研究是 LSES-Ⅲ阶段中的类型Ⅰ混合设计的一个典型案例。在这个混合设计案例中，确证性研究的作者们使用了本书第5章提到的定量化方法论。课堂观察工具（Classroom Observation Instrument，COI）所获得的开放型资料被转化为数值数据，这些数据随后又被研究者应用 t-测试和 MANOVAs/ANOVAs 加以分析。研究者试图判断不同类型的课堂教学行为是否发生于不同效率的学校。他们预估，较之于无效率学校的教师，更具效率的学校的特征在于其教师能提供更高品质的教学。教师的品质是通过 Stallings 的课堂快照法（Classroom Snapshot）完成测量（Stallings，1980；Stallings and Kaskowitz，1974）的，由此产生出关于作业时间和 COI 的数据指数。

研究者们发展出 COI，旨在以田野笔记的形式对 15 个有关教师效率的一般指标提供高度推论性的课堂资料。这些指标建立在从 Rosenshine（1983）以及 Rosenshine 和 Stevens（1986）评论中所收集到的关于有效教学的各个特征之上，包括对作业时间、最初的学生练习、新资料的展示和积极强化等内容的测量。与“最初的学生练习”相关的具体提示包括提问的高频率、教师引导的交流、教师的鼓励、所有学生回答问题的机会以及在最初学生练习中 80% 的成功率。每当在课堂中观察到与某种类型相关的行为出现时，观察者就会随之进行记录，从而对上述 15 个指标逐一完成田野笔记。

在 LSES-Ⅲ阶段，观察者将 700 个课堂内的教师和学生的行为进行了编码处理。总计有超过 2 500 个信息单位（700 个课堂 × 15 个 COI 观察项目 ×2.5 个（平均对每个项目的）观察员评论）被记录下来。为了减少研究资料的总量，研究者作了定量化处理。COI 田野笔记由独立的评估者进行分析。根据对每一个受到观察的教师的所有记录，两名评估者对有效教学的每个指标进行打分，分值为从 1（有效教学的充分证据）、2（有效教学的不完全证据）到 3（有效教学缺乏证据）。统计分析的结果表明，在更具效率的学校里的教师的教学质量的确要比那些在无效率的学校里的

教师的教学质量更高。

LSES 的这一部分研究工作属于类型 I 混合模型设计。这一研究是确证性的,其预先提出了假设,认为有效率学校与无效率学校的不同点在于前者的教师所提供的教学品质要高于后者的教师。资料的收集和操作是纯定性的,观察者对 COI 的应答产生了大量的田野笔记形式的叙述性资料;分析过程是定量的,研究者使用标准的统计方法对已经用定量化技术转化为数值的研究资料进行分析。

类型Ⅱ 混合设计:
确证性研究、定性资料/操作、定性分析和推论

Teddlie 等人(1989)研究的另一方面包括使用案例研究路径,以进一步理解为什么更有效率的教学进程发生在更有效率的学校,而效率低的教学进行发生在效率低的学校。研究的一个发现是,相较于无效率的学校里的教师,更有效率的学校里的教师在行为上表现得更为一致(也就是说,他们行为的标准差要小得多)。检验定性资料的目的之一是弄清教师行为的差异(即平均差与方差)是如何发生的。

LSES-Ⅲ阶段的定性资料包括对学校和课堂进行观察时所记录下来的田野笔记,以及对校长和教师所进行的访谈。研究者抽取出一组学校进行配对,在这个配对组中的两所学校的教学品质之间差异很大。研究者随后仔细审读关于这两个学校的田野笔记,试图由此确定期间所发生的不同进程。他们的定性描述包含在 Teddlie 等人(1989)所著一书的结论部分,并且聚焦在效率不同的学校的校长身上:其中一个校长在她的学校里是学习时间的严格"守卫者",而另一个校长则允许教员们对学习时间得过且过。

"路易斯安那学校效率研究(LSES)"的这一部分研究工作属于确证性研究,研究者预先设计了假设,认为高效率学校的特点在于其教师能够比无效率学校的教师提供更高品质的教学。研究资料的收集和操作是纯定性的,包括对学校和课堂进行观察时所记录下来的田野笔记,以及对校长和教师所进行的访谈。分析和推论过程也是定性的,研究者使用了归纳取向的方法,研究者

依此对这些田野资料进行仔细研究，以得到生成性主题，用以描述此前已发现的统计差异之背后的事态进程。

类型Ⅲ 混合设计：
探索性研究、定量数据/操作、统计分析和推论

在前面提到的使用纯定性设计的（Stringfield and Teddlie, 1990）研究工作中，也使用了一种类型Ⅲ的混合设计，这是在《学校效率及学校的改进》一书的编辑的坚持下所完成的。编辑要求研究者在发现纯定性研究的结论之前看一看定量的数据并对之进行分析。这些编辑希望确保研究者观察到的是“真实”的学校进步，他们坚持研究者要为学校进步这一效果提供“过硬”的资料（即表明学校在标准化测试数据方面有相比以前更高的得分）。

因此，研究者使用 1982—1984、1984—1985、1989—1990 学年的信息，对此前各学校的 CRT 和 NRT 数据进行了检验。从这些定量资料中所得到的结论与此前收集的定性资料正相吻合，于是确证了所研究的 4 所学校中正发生某些方面的改进。考虑到事先没有设立关于“自然进步”学校的改进性假设，LSES 的这一部分工作是探索性的，研究中的数据收集和操作是定量的，包括研究者所做的 NRT 数据和国家管理的 CRT 数据。资料分析在本质上属于统计分析，研究者把从这 4 所进步中的学校所获得的数据，与此前有关它们的测试结果以及从它们的配对学校中所得到的结果进行了比较。

类型Ⅳ 混合设计：
探索性研究、定性资料/操作、统计分析和推论

Crone 和 Teddlie（1995）对于不同效率学校里的教师行为的研究，在一定程度上重复了前人的研究（如 Teddlie, et al., 1989; Virgilio, Teddlie and Oescher, 1991）。在他们研究工作中的访谈部分，研究者针对在有效和无效学校的教师行为中所发现的不同结果模式，向接受访谈的教师和校长追问了有关原因：在更有效率的学校里，有效教学行为的比例更高，并且对有效行为的离散值也更小；而在无效率的学校中，有效教学行为的比例更低，并且对无效行为的离散值也更大。

研究者使用了一种“访谈引导方法”(Patton,1990),预先以大纲的形式具体指定了访谈的主题和问题,使访问者在访谈进行过程中能确定问题的准确顺序和措辞。访谈的主题相当宽泛,涉及教师的选拔和社会化、教员之间的凝聚力、校长所实施的变革等。

研究者在研究工作的访谈部分并未提前预设任何假设,他们将这一部分视为探索性的,用来判定不同效率的学校里的教师和校长对其经历(如教师的选拔和社会化)的看法有何不同。对访谈引导中开放性问题的回答,成为研究者叙述性田野笔记的来源,并成为相关研究资料。

在研究者开始使用 Lioncoln 和 Guba(1985)所描述的常数对比方法对资料进行分析时,许多主题开始生成。这些主题出现的频率高低看来似乎取决于访谈者究竟是在高效率学校还是在无效率学校进行的访谈。研究者决定对这些不同主题的数量进行计算,然后将所获得的数据资料进行卡方分析。

这种分析方法属于本书第 6 章所讨论的定量化方法论的又一个典型案例。对访谈的分析表明:

> 在有效率的学校里,教师们之间的凝聚力更强……访谈包括了 13 个高效率学校的教师和 14 个无效率学校的教师……调查者向所有教师询问了他们对学校的目标有什么促进办法。高效率学校有 12 个教师的回答是教师们一起努力,而无效率学校却只有两个教师答复是教师们为目标共同努力(卡方 =7.17,$p < 0.01$)。(Crone and Teddlie,1995:5)

Crone 和 Teddlie(1995)的这一部分研究工作属于类型Ⅳ混合设计。研究是探索性的,预先没有进行建立假设;资料收集和操作是定性的,研究者使用访谈产生叙述性田野笔记;对资料的分析和推论过程是定量的,他们使用卡方分析对已经实现叙述性资料定量化的频率计算数据进行分析。

类型Ⅴ 混合设计:

确证性研究、定量数据/操作、定性分析和推论

在表 3.1 中,类型Ⅳ混合设计和类型Ⅴ混合设计被描述成“相当少见”,其主要原因在于它们涉及本书第 6 章所讨论的定性化方法论:将定量数据转化叙述性数据,然后进行定性分析。许

多行为和社会科学的研究者会在上述类型Ⅰ和类型Ⅳ混合模型研究中使用更普遍的定量化工具，但却很少有研究者尝试使用定性化方法来形成形描和类型范畴。

在确证性研究中使用这种定性化工具进行研究的最新案例是 Durland 和 Teddlie 所作的研究（Durland, 1996; Durland and Teddlie,1996）。这项研究要对 LSES-Ⅵ中的社会人际关系测量资料进行分析。社会人际关系测量问卷产生了一些本质上属于混合性的资料：以社会关系测量形式提供的定量数据（数字指数）和以社会关系图形形式提供的定性资料（依据社会单位间关系而绘制的二维度图式）。

社会关系测量问题通常十分简单。例如，在 LSES-Ⅵ阶段，主要的问题是要求教员指出并排列出他们在过去两周内交流学术问题最多的 3 个人。通过这个简单的问题，可以产生出非常复杂的定量的社会测量数据（如对校长的向心度的数值测量，对教员之间的凝聚力和密度的数值测量，等等）以及定性的社会关系网图（关于教职工交往模式的图示，其中包括小集团、彼此孤立等）。在 LSES-Ⅵ阶段，研究者同时使用了社会关系测量法和社会关系图式，但是可以肯定的是，社会关系图式所产生的信息是最有价值的。社会关系图式使研究者可以根据现存的学校效率研究，对关于有效率学校与无效率学校中应当具备的教师互动模式的预测进行确证。

Durland（1996）提出了一个“关于学校效率的向心度—凝聚力模型”，这一模型预测了她所设想的学校中存在的社会关系图式，这些图式的具体类型会依据学校效率指数（SEI，有效率或无效率）以及学生群体的社会经济地位（中等水平或低等水平）的不同而发生变化。这些预测性的社会关系图式或教员互动样式的可视表达，立基于 SER 理论和关于在这些学校有望发现的领导类型和交往模式的研究之上。这项研究使用了一种先进的计算机程序——KrackPlot（Krachhardt, Lundberg and O'Rourke, 1993），以根据经验数据产生出相应的社会关系图式。随后，研究者可以比较研究从经验中推出的社会关系图式与此前依据预设理论所推导出的社会关系图式。

社会关系图式的分析显示：①高效率学校很可能是更稠密

的、更具向心集中形状的("完全网状的"),而无效率学校则可能是更发散的和线性的("纤维状的");②较之于无效率学校,高效率学校的领导或代理领导在高效率学校中更有可能处于社会关系图式的中心位置,在无效率学校中领导则往往处于边缘位置;③无效率学校有着更多的孤立群体和小集团,而高效率学校则相反。

LSES-Ⅵ研究的这一部分工作属于确证性研究,研究者在现存 SER 文献的基础上预先建构了假说,预测教员交往模式存在着各种类型。资料的收集和操作是定量的,研究者使用社会关系测量的调查问卷,要求教员根据与之交往的程度对学校里的其他个人进行排序;为这项研究的描述文字所作的分析和推论则是定性的,研究者使用特定的计算机程序(KrackPlot),根据经验的数值数据生成关于高效率学校和无效率学校的教员交往模式的图形。

类型Ⅵ 混合设计:
探索性研究、定量数据/操作、定性分析和推论

LSES-Ⅳ研究也产生了一个使用定性化工具进行探索性研究的典型案例。如表 8.1 所示,LSES-Ⅳ对 8 对学校的配对组进行了长达 5 年的跟踪调查。根据 1982—1983 学年以及 1983—1984 学年所获得的基线数据,研究者将这些学校最初分类为高效率学校和无效率学校两组:它们的效率状况是通过 LSES-Ⅲ阶段 1984—1985 学年所进行的实地访问而识别的。尽管所有学校在 LSES-Ⅲ阶段都保持了它们最初的分类地位,但也有一些证据显示,有一些自然发生的学校进步——这在前面已经描述过了。

研究者期望,到 LSES-Ⅳ阶段的时候,这 16 所学校当中的一些会在其效率状况上发生不同的变化,但他们决定不对这些变化提前作出预测。事实上,研究者甚至不能确定,到 1989—1990 学年(即 LSES-Ⅳ阶段),最初的二分法分类框架(即高效率与无效率)是否还足以刻画这些已被列为研究对象的学校。

研究者并没有去预测这些学校在 LSES-Ⅳ阶段时的状态,他们决定使用一套多维度的定量标准来确定学校的效率状态,以及现实的效率类型。Teddlie 和 Stringfield(1993:84-85)为每个学校的 8 个定量标准提供了数据:标准参考测试的分数、规范参考测试

的分数、课堂教学质量的高推论性测量分数、课堂内的作业时间、教员的稳定性、研究者对整个学校氛围测试的分值（学校教育的等级维度）以及学校学生群体的社会经济地位的变化。

在这些定量数据的基础上，研究者提出了一个新的分类框架，将这些学校分成4种定性界定的类型：稳定的更具高效率的学校、进步中的学校、退步中的学校和稳定的无效率的学校。LSES-Ⅳ研究的这一部分工作是探索性的，研究者预先没有对将会出现的效率类型建立任何假设。资料收集和操作是定量的，研究者使用了8个独立的关于效率的数值指数；分析和推论是定性的，研究者使用了定性化工具将数值数据转化为关于16个学校的叙述性资料（案例研究），然后对之进行定性分析，产生出4种生成的类型范畴。

9

结论与展望

本书的主要目的是将正在建设中的横跨定性和定量两种研究传统的哲学和方法论的“桥梁”加以延展。这两类研究方法都有丰富的传统，都有热情的拥趸，其中一些人在过去的20年里花费了巨大的精力将这两种方法分离开来。因此，我们将这两种方法调和起来的努力是颇具挑战压力的，但同时也是大有裨益的。我们不得不广博地涉猎从抽象的哲学基础到现实的田野研究中“亲身体验的案例”等多个不同领域内的材料。我们希望能够证明，将定量研究方法与定性研究方法结合起来不仅是可能的，而且在许多不同的研究背景中也是非常有益的。这一点对研究生尤其重要，他们经常对在其学位论文的研究中使用混合方法缺乏信心。

我们在本书当中自始至终强调的第二个要点是，相对于有关研究方法或研究范式的考虑，研究问题本身才具有更重要的地位。我们在本书前面的“问题为王”的部分已经对此作过讨论。研究的问题决定着研究的设计、资料收集的方法等。最好的研究方法就是在解决研究问题上最有效的，并且推论质量最好（信度和内部效度）的方法。混合方法常常比单独使用定性路径或定量路径能更有效地解决研究问题。我们经常会听到研究生说：“如果我打算用定性方法进行我的学位论文研究，那我为什么要学习定量方法和统计分析?”喜欢定量方法的学生也会提出相似的问题。通过强调研究问题的重要性和混合方法对这些问题给出综合解答的优越性，我们希望本书为这两种学生都能提供某些启示。

本书的结论之一是，我们可以将实用主义的范式作为使用混合方法和混合模型的哲学基础，尤其是在认识论、价值论和存在论方面。实用主义者拒绝范式之争所带来的非此即彼的观点。

我们还强调，对混合方法和混合模型进行分类是必要的，因为当前的研究文献存在着概念混淆不清的问题。因此，我们首先提出了一个分类法，将其分成单一方法、混合方法和混合模型研究。然后，我们对建立在三个基本的维度、步骤或研究过程之上的混合模型设计进行了分类，即研究的类型（探索型对确证型）、资料收集和操作的类型（定性方法对定量方法）以及分析和推论的类型（定性方法对统计方法）。这些分类是暂时性的，而我们提出这些分类是为了激发方法论研究者对归类混合方法和混合模型研究的最佳方式进行切磋。

在有限的篇幅之内，我们试图证明，社会和行为科学领域内越来越多的研究者正在使用 8 种混合模型设计（其中 6 种比较简单，另外 2 种比较复杂）。在这 8 种混合模型设计中，研究过程的所有阶段都存在定性路径和定量路径的混合。在很多研究案例之中，研究者在运用这些混合模型时都没有意识到它们是混合性质的。我们从这些案例总结出，研究老手使用复杂的混合模型（类型Ⅶ和类型Ⅷ）看上去比使用那些简单的混合模型（类型Ⅰ—Ⅵ）更多，而对于研究新手来说，从使用更为简单的混合模型开始可能会更好。

我们试图在定性路径和定量路径之间就推论质量（内部效度/信度）的问题架起一座桥梁。由于这个问题对所有的研究者都具有重要的意义，我们深信，对推论的内部效度或信度的评估是对定性和定量两种路径进行调和的一个关键问题。作为教授研究方法论的一个组成部分，我们经常要求学生从不同的学科角度对研究报告进行批判性评估。他们几乎每次都会意识到，最重要的问题是在作出结论时对结果的阐释。我们相信许多同事也从他们的学生那里得到了相似的反馈。

这一问题并非定性路径或定量路径所独有的。考察一下各种期刊论文的结论部分，可以极其容易地证明这一点。在定性研究传统和定量研究传统中，都有一些研究者不恰当地作出了超过其资料所允许的通则化，他们所表达的结论：①不具内部效度，因

为它们所阐释的关系可以有不同的解释;②不可相信,因为并未建立充分可信的工具(如长时间的投入、持续的观察和成员互检)。

我们认为,一个好的研究报告应当能使读者得出具有一定确然程度的结论。研究者的职责就是在根据结果作出推论时,提供可以达到这种确然性的手段。研究者对所研究的变量间关系进行推论的质量,取决于研究者和已了解结果的读者在多大程度上能够确信,除了他们所提出的推论,不存在任何其他合理的解释。

对于这两种研究路径,都存在着一些怀疑和误解。许多定量研究之所以受到批评是因为它强调统计显著性(偶然获得的某一关系的概率),却牺牲了对关系/效果之量级的关注。批评者还指责定量研究不能在分析中吸纳更广泛的信息,如观察和非结构化的应答。甚至当这些资料收集起来以后,它们也往往被定量研究通过定量化降解为简单的类型范畴。另一方面,定性研究者所经常遇到的批评则是他们在报告结果时具有选择性(这一事实的提示在于,根本没有什么可用来提出报告的简单资料之简化方法),因为其没有提供“用以增强发现之信度的替代性视角”(Murphy and O’Leary,1994:214),也因为其所表达的是个人观点而非准确地反映其信息提供者的视角。

许多这样的误解(但有时也是正确诊断的弊病!)可以轻松地得到缓解,办法就是使用混合研究方法,以及对通过这些方法所获得的发现的内部效度和信度进行严格的评估。三角测量工具涉及对定性和定量的资料来源的调和,这为改进推论的质量提供了一剂灵丹妙药。

参考文献

Aghajanian, A., & Moghadas, A. A. (in press). Trends, correlates and consequences of divorce in Iran. *Journal of Divorce and Remarriage*, 29.

Ary, D., Jacobs, L. C., & Razavieh, A. (1996). *Introduction to research in education* (5th ed.). Fort Worth: Harcourt Brace College.

Babbie, E. (1992). *The practice of social research*. Belmont, CA: Wadsworth.

Bacharach, S., Bauer, S. C., & Shedd, J. B. (1986). *The learning workplace: The conditions and resources of teaching*. (ERIC Document Reproduction Service No. ED 279 614)

Bannister, R. C. (1987). *Sociology and scientism: The American quest for objectivity*, 1890-1940. Berkeley: University of California Press.

Barnett, D. W., & Macmann, G. M. (1990). Personality assessment: Critical issues for research and practice. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (Vol. 2). New York: Guilford.

Berelson, B. (1952). *Content analysis in communication research*. Glencoe, IL: Free Press.

Bernstein, R. (1983). *Beyond objectivism and relativism*. Philadelphia: University of Pennsylvania Press.

Blalock, H. M. (1978). Ordering diversity. *Society*, 15, 20-22.

Bogdan, R. C., & Biklen, S. J. (1982). *Qualitative research for education: An introduction to theory and method*. Boston: Allyn & Bacon.

Brewer, J., & Hunter, A. (1989). *Multimethod research: A synthesis of styles*. Newbury Park, CA: Sage.

Brien, T. M., Dennis, L. T., Mahoney, C. A., & Wallnau, L. (1994). Dimensions of self-efficacy among three groups of condom users. *Journal of American College Health*, 42, 167-174.

Brookover, W, B., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *Schools, social systems and student achievement: Schools can make a difference*. New York: Praeger.

Brown, S., Riddell, S., & Duffield, J. (1996). Possibilities and problems of small-scale studies to unpack the findings of large-scale studies of school effectiveness. In J. Gray, D. Reynolds, C. Fitz-Gibbon, & D. Jesson (Eds.), *Merging traditions: The future of research on school effectiveness and school improvement* (pp. 93-120). London: Cassell.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Campbell, D., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 54, 297-312.

Campbell, D. T., & Stanley, J. (1966). *Experimental and quasi-experimental design for research*. Chicago: Rand McNally.

Caracelli, V. W., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15(2), 195-207.

Carey, J. W. (1993). Linking qualitative and quantitative methods: Integrating cultural factors into public health. *Qualitative Health Research*, 3, 298-318.

Chandler, L. A. (1990). The projective hypothesis and the development of projective techniques for children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (Vol. 2). New York: Guilford.

Cherryholmes, C. C. (1992). Notes on pragmatism and scientific realism. *Educational Researcher*, 21, 13-17.

Chrispeels, J. H. (1992). *Purposeful restructuring: Creating a culture for learning and achievement in elementary schools*. London: Falmer.

Cook, T. D., & Campbell, D. T. (1979). *Quasiexperimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Creswell, J. W. (1995). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Crone, L., & Teddlie, C. (1995). Further examination of teacher behavior in differentially effective schools: Selection and socialization processes. *Journal of Classroom Interaction*, 30(1), 1-9.

Datta, L. (1994). Paradigm wars: A basis for peaceful coexistence and beyond. In C. S. Reichardt & S. F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 53-70). San Francisco: Jossey-Bass.

Davidson, D. (1973). On the very idea of a conceptual scheme. *Proceedings of the American Philosophical Association*, 68, 5-20.

Denzin, N. K. (1978). The logic of naturalistic inquiry. In N. K. Denzin (Ed.), *Sociological methods: A sourcebook*. New York: McGraw-Hill.

Denzin, N. K. (1992). *Symbolic interactionism and cultural studies*. Cambridge: Basil Blackwell.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.

DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.

Dooley, D. (1995). *Social research methods* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Dressler, W. W. (1991). *Stress and adaptation in the context of culture: Depression in a southern black community*. Albany: State University of New York Press.

Durland, M. (1996). *The application of network analysis to the study of differentially effective schools*. Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.

Durland, M., & Teddlie, C. (1996, April). *A network analysis of the structural dimensions of principal leadership in differentially effective schools*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Duval, S., & Wicklund, R. A. (1972). *A theory of objective self-awareness*. New York: Academic Press.

Eisner, E. W. (1975). *The perceptive eye: Toward the formulation of educational evaluation* (Occasional Papers of the Stanford Evaluation Consortium). Stanford, CA: Stanford University, Press.

Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York: Macmillan.

Fals-Stewart, W., Birchler, G., Schafer, J., & Lucente, S. (1994). The personality of marital distress: An empirical typology. *Journal of Personality Assessment*, 62, 223-241.

Fetterman, D. M. (1992). In response to Lee Sechrest's 1991 AEA presidential address: "Roots: Back to our first generations," February 1991, 1-7. *Evaluation Practice*, 13, 171-172.

Floyd, J. A. (1993). The use of across-method triangulation in the study of sleep concerns in healthy older adults. *Advances-in-Nursing-Science*, 16(2), 70-80.

Fowler, F. J. (1993). *Survey research methods*. Newbury Park, CA: Sage.

Fowler, F. J., & Mangione, T. W. (1990). *Standardizing interviewer-related error*. Newbury Park, CA: Sage.

Franklin, R. D., Allison, D. B., & Gorman, B. S. (1996). *Design and analysis of single case research*. Mahwah, NJ: Lawrence Erlbaum.

Freeman, D. (1983). *Margaret Mead and Samoa: The making and unmaking of an anthropological myth*. Cambridge, MA: Harvard University Press.

Freeman, J., & Teddlie, C. (1996, April). *A phenomenological examination of "naturally occurring" school improvement: Implications for democratization of schools*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Gage, N. (1989). The paradigm wars and their aftermath: A "historical" sketch of research and teaching since 1989. *Educational Researcher*, 18, 4-10.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). New York: Longman.

Gardner, M. (1993). The great Samoan hoax. *Skeptical Inquirer*, 17, 131-135.

Geertz, C. (1973). *The interpretation of culture*. New York: Basic Books.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, 26, 309-320.

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, 40, 266-275.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in educational research*. New York: Academic Press.

Goodman, N. (1984). *Of mind and other matters*. Cambridge, MA: Harvard University Press.

Gravetter, F. J., & Wallnau, L. B. (1998). *Statistics for the behavioral sciences*. St. Paul, MN: West.

Greenberg, J., & Folger, R. (1988). *Controversial issues in social research methods*. New York: Springer-Verlag.

Greene, J. C. (1994). Qualitative program evaluation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 530-544). Thousand Oaks, CA: Sage.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11, 255-274.

Guba, E. G. (1987). What have we learned about naturalistic evaluation? *Evaluation Practice*, 8, 23-43.

Guba, E. G. (1990). *The paradigm dialog*. Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (1990). *Fourth-generation evaluation*. Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.

Haak, R. A. (1990). Using sentence completion to assess emotional disturbance. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (Vol. 2). New York: Guilford.

Hallinger, P., & Murphy, J. (1986). The social context of effective schools. *American Journal of Education*, 94, 328-355.

Hammersley, M. (1989). *The dilemma of qualitative method: Herbert Blumer and the Chicago tradition*. London: Routledge Kegan Paul.

Hammersley, M. (1992). *What's wrong with ethnography*. London: Routledge Kegan Paul.

Hanson, N. R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge: Cambridge University Press.

Harris, R. (1985). *A primer of multivariate statistics*. Orlando, FL: Academic Press.

Hedrick, T. E., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide*. Newbury Park, CA: Sage.

Henry, G. T. (1990). *Practical sampling*. Newbury Park, CA: Sage.

Hooper, M. L. (1994). The effects of high and low level cognitive and literacy language arts tasks on motivation and learning in multiability, multicultural classrooms. *Developmental Studies: Learning-and-Instruction*, 4(3), 233-251.

House, E. R. (1994). Integrating the quantitative and qualitative. In C. S. Reichardt & S. F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 13-22). San Francisco: Jossey-Bass.

Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher*, 17, 10-16.

Huberman, A. M., & Miles, M. B. (1994). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428-444). Thousand Oaks, CA: Sage.

Hughes, J. N., Barker, D., Kemenoff, S., & Hart, M. (1996). Problem ownership, causal attributions, and self-efficacy as predictors of teachers' referral decisions. *Journal of Educational and Psychological Consultation*, 4, 369-384.

Ingles, S. J., Scott, L. A., Lindmark, J. T., Frankel, M. R., Myers, S. L., & Wu, S. (1992). *National Education Longitudinal Study of* 1988: *First follow-up: Student component data file user manual* (Vol. 1). Washington, DC: U. S. Department of Education.

Iwanicki, E. E., & Tashakkori, A. (1994). *The proficiencies of the effective principal: A validation study*. Baton Rouge: Louisiana Department of Education.

Janesick, V. J. (1994). The dance of qualitative research design: Metaphor, methodolatry, and meaning. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 209-219). Thousand Oaks, CA: Sage.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24, 602-611.

Jimenez, R. T., Garcia, G. E., & Pearson, P. D. (1995). Three children, two languages, and strategic learning: Case studies in bilingual/monolingual reading. *American Educational Research Journal*, 32, 67-98.

Joffe, A., & Radius, S. M. (1993). Self-efficacy and intent to use condoms among entering college freshmen. *Journal of Adolescent Health*, 14, 262-268.

Johnson, M. J., & Pajares, F. (1996). When shared decision making works: A 3-year longitudinal study. *American Educational Research Journal*, 33, 599-627.

Jones, E. E., & Davis, K. E. (1966). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Freeman.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, E. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attributions: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.

Keirsey, D., & Bates, M. (1984). *Please understand me: Character and temperament types*. Del Mar, CA: Prometheus Nemesis.

Kerlinger, F. N. (1986). *Foundations of behavioral research*. New York: Holt, Rinehart & Winston.

Kidder, L. H. (1981). Qualitative research and quasi-experimental frameworks. In M. B. Brewer & B. E. Collins (Eds.), *Scientific inquiry and the social sciences*. San Francisco: Jossey-Bass.

Kneller, G. F. (1984). *Movements of thought in modern education*, New York: John Wiley.

Knoff, H. M. (1990). Evaluation of projective drawings. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (Vol. 2). New York: Guilford.

Kochan, S., Tashakkori, A., & Teddlie, C. (1996, April). *You can't judge a high school by achievement alone: Preliminary findings from the construction of behavioral indicators of school effectiveness*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Krackhardt, D., Lundberg, M., & O'Rourke, L. (1993). KrackPlot: A picture's worth a thousand words. *Connections*, 16, 37-47.

Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach.* White Plains, NY: Longman.

Krueger, R. A. (1988). *Focus groups: A practical guide for applied research.* Newbury Park, CA: Sage.

Kuhn, T, S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Latane, B., & Rodin, J. (1969). A lady in distress: Inhibiting effects of friends and strangers on bystander intervention. *Journal of Experimental Social Psychology*, 5, 189-202.

Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, and supervision.* Newbury Park, CA: Sage.

LeCompte, M. D., & Preissle, J., with Tesch, R. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). New York: Academic Press.

Leininger, M. M. (1985). *Qualitative research methods in nursing.* New York: Grune & Stratton.

Lincoln, Y. S. (1991). The arts and sciences of program evaluation. *Evaluation Practice*, 12, 1-7.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Mangione, T. W. (1995). *Mail surveys: Improving the quality.* Thousand Oaks, CA: Sage.

Mann, A. F. (1994). College peer tutoring: Maps of development. *Journal of College Student Development*, 35, 164-169.

Manning, P. K., & Cullum-Swan, B. (1994). Narrative, content, and semiotic analysis. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 463-477). Thousand Oaks, CA: Sage.

Marin, G., & Matin, V. M. (1991). *Research with Hispanic populations.* Newbury Park, CA: Sage.

Marsh, H. W. (1994). Using the National Longitudinal Study of 1988 to evaluate theoretical models of self concept: The Self-Description Questionnaire. *Journal of Educational Psychology*, 86, 439-456.

Maruyama, G., & Deno, S. (1992), *Research in educational settings.* Newbury Park, CA: Sage.

Maxwell, J. A. (1996). *Qualitative research design: An interpretive approach.* Thousand Oaks, CA: Sage.

McMillan, J. H. (1996). *Educational research: Fundamentals for the consumer* (2nd ed.). New York: HarperCollins.

McMillan, J. H., & Schumacher, S. (1997). *Research in education.* New York: Longman.

McNamara, J. F. (1994). *Surveys and experiments in education research.* Basel, Switzerland: Technomic.

Meekers, D. (1994). Combining ethnographic and survey methods: A study of the nuptiality patterns of the Shona of Zimbabwe. *Journal of Comparative Family Studies*, 25(3), 313-328.

Miles, M., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed,). Thousand Oaks, CA: Sage.

Miller, W. L., & Crabtree, B. J. (1994). Clinical research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 340-352). Thousand Oaks, CA: Sage.

Mishler, E. G. (1990). Validation in inquiry-guided research: The role of examples in narrative studies. *Harvard Educational Review*, 60, 415-441.

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40, 120-123.

Murphy, C. M., & O'Leary, K. D. (1994). Research paradigms, values, and spouse abuse. *Journal of Interpersonal Violence*, 9, 207-223.

Murphy, J. P. (1990). *Pragmatism: From Peirce to Davidson.* Boulder, CO: Westview.

Nielsen, K. (1991). *After the demise of the tradition: Rorty, critical theory, and the fate of philosophy.* Boulder, CO: Westview.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Pemberton, M. B., Insko, C. A., & Schopler, J. (1996). Memory for and experience of differential competitive behavior of individuals and groups. *Journal of Personality and Social Psychology*, 71, 953-966.

Phillips, D. C. (1990). Postpositivist science: Myths and realities. In E. Guba (Ed.), *The paradigm dialog.* Newbury Park, CA: Sage.

Popper, K. R. (1959). *The logic of scientific discovery.* New York: Basic Books.

Reichardt, C. S., & Rallis, S. F. (1994). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. In C. S. Reichardt & S. F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 85-92). San Francisco: Jossey-Bass.

Reis, H. R., & Wheeler, L. (1991), Studying social interaction with the Rochester interaction record. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 269-318). New York: Academic Press.

Richardson, L. (1994). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 209-219). Thousand Oaks, CA: Sage.

Roberts, P., & Le-Dorze, G. (1994). Semantic verbal fluency in aphasia: A quantitative and qualitative study in test-retest conditions. *Aphasiology*, 8(6), 569-582.

Robinson, J. P., & Shaver, P. R. (1975). *Measures of social psychological attitudes.* Ann Arbor, MI: Institute for Social Research.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes.* San Diego, CA: Academic Press.

Rorty, R. (1982). Pragmatism, relativism, and irrationalism. In R. Rorty (Ed.), *Consequences of pragmatism* (pp. 160-175). Minneapolis: University of Minnesota Press.

Rorty, R. (1990). Introduction. In J. P. Murphy (Ed.), *Pragmatism: From Peirce to Davidson.* Boulder, CO: Westview.

Rosenberg, M. (1979). *Conceiving the self.* New York: Basic Books.

Rosenshine, B. (1983). Teaching functions in instructional programs. *Elementary School Journal*, 83, 335-351.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. Wittrock (Ed.), *Third handbook of research on teaching.* New York: Macmillan.

Rosenthal, R. (1976). *Experimenter effects in behavioral research* (enlarged ed.). New York: Irvington.

Rosenthal, R. (1991). *Meta-analysis procedures for social research.* Newbury Park, CA: Sage.

Rosnow, R. R., & Rosenthal, R. (1996). *Beginning behavioral research: A conceptual primer* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Ross, D. (1991). *The origins of American social science.* Cambridge: Cambridge University Press.

Rossi, P. H. (1994). The war between the quals and quants: Is a lasting peace possible? In C. S. Reichardt & S. F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 23-36). San Francisco: Jossey-Bass.

Rusbult, C. E., Onizuka, R. K., & Lipkus, I. (1993). What do we really want? Mental models of ideal romantic involvement explored through multidimensional scaling. *Journal of Experimental Social Psychology*, 29, 493-527.

Salkind, N. J. (1997). *Exploring research.* Upper Saddle River, NJ: Prentice Hall.

Schuyten, S. (1995, April). *The effects of assessor and assessee gender, ethnicity, and assessor's role on performance assessment of teachers.* Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.

Schuyten, S., & Tashakkori, A. (1995). *Effects of assessor and assessee gender and ethnicity, and assessor's role on performance assessment of teachers.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 118-137). Thousand Oaks, CA: Sage.

Sechrest, L. (1991). Roots: Back to our first generations. *Evaluation Practice*, 13, 1-7.

Sechrest, L., & Sidani, S. (1995). Quantitative and qualitative methods: Is there an alternative? *Evaluation and Program Planning*, 18(1), 77-87.

Shadish, W. R. (1995). Philosophy of science and the quantitative-qualitative debates: Thirteen common errors. *Evaluation and Program Planning*, 18, 63-75.

Sinclair, A. (1994). Prediction-making as an instructional strategy: Implications of teacher effects on learning, attitude toward science, and classroom practices, *Journal of Research and Development in Education*, 27, 153-161.

Smith, J. K. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. *Educational Researcher*, 12. 6-13.

Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational researchers. *Educational Researcher*, 15, 4-12.

Smith, M. L. (1994). Qualitative plus/versus quantitative: The last word. In C. S. Reichardt & S. F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 37-44). San Francisco: Jossey-Bass.

Smith, T. E., Sells, S. P., & Clevenger, T. (1994). Ethnographic content analysis of couple and therapist perceptions in a reflecting team setting. *Journal of Marital and Family Therapy*, 20(3), 267-286.

Spradley, J. P. (1979). *The ethnographic interview.* New York: Holt, Rinehart & Winston.

Spradley, J. P. (1980). *Participant observation.* New York: Holt, Rinehart & Winston.

Stallings, J. A. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher*, 9(11), 11-16.

Stallings, J. A., & Kaskowitz, D. (1974). *Follow through classroom observation evaluation* (1972-1973). Menlo Park, CA: SRT International.

Stange, K. C., & Zyzanski, S. J. (1989). Integrating qualitative and quantitative research methods. *Family Medicine*, 21, 448-451.

Stevens, J. P. (1996). *Applied multivariate statistics for the social sciences.* Mahwah, NJ: Lawrence Erlbaum.

Stewardt, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice.* Newbury Park, CA: Sage.

Stewart, D. W., & Kamins, M. A. (1993). *Secondary research: Information sources and methods* (2nd ed.), Newbury Park, CA: Sage.

Stringfield, S. (1994). Outlier studies of school effectiveness. In D. Reynolds, B. Creemers, P. Nesserlrodt, E. Schaffer, S. Stringfield, & C. Teddlie, *Advances in school effectiveness research and practice.* Oxford: Pergamon.

Stringfield, S., & Teddlie, C. (1990). School improvement efforts: Qualitative and quantitative data from four naturally occurring experiments in Phases Ⅲ and Ⅳ of the Louisiana School Effectiveness Study. *School Effectiveness and School Improvement*, 1(2), 139-162,

Stringfield, S., & Teddlie, C. (1991). Observers as predictors of schools' effectiveness status. *Elementary School Journal*, 91(4), 357-376,

Takane, Y., Young, F. W., & DeLeeuw, J. (1977). Nonmetric individual difference multidimensional scaling: An

alternative least squares method with optimal scaling features. *Psychometrika*,42,7-67.

Tashakkori,A. (1993). Gender,ethnicity,and the structure of self-esteem: An attitude theory approach. *Journal of Social Psychology*,4,479-488.

Tashakkori,A.,Aghajanian,A.,& Mehryar,A. H. (1996,July). *Consistency of Iranian adolescent behavioral intentions across two decades of change*. Paper presented at the 54th Annual Convention of the International Council of Psychologists,Banf,Alberta,Canada.

Tashakkori,A.,Boyd,R.,& Sines,M. (1996,July). *Predictors of drop-out and persistence of 8th grade Hispanic youth in the U. S.* Paper presented at the 54th Annual Convention of the International Council of Psychologists,Banf,Alberta,Canada.

Tashakkori,A.,Haghighat,S.,& Yousefi,F. (1990). Effects of preschool education on intelligence and achievement of a group of Iranian elementary school children. *International Review of Education*,34,499-508.

Tashakkori,A.,& Thompson,V. D. (1991). Social change and change in intentions of Iranian youth regarding education, marriage,and careers. *International Journal of Psychology*,26,885-893.

Tashakkori,A.,Thompson,V,D.,& Mehryar,A. H. (1987). Iranian adolescents' intended age of marriage and desired family size. *Journal of Marriage and the Family*,49,917-924.

Taylor,D. L.,& Tashakkori,A. (1995). Participation in decision making and school climate as predictors of teachers' job satisfaction and sense of efficacy. *Journal of Experimental Education*,63,217-230.

Taylor,D. L.,& Tashakkori,A. (1997). Toward an understanding of teachers' desire for participation in decision making. *Journal of School Leadership*,7,1-20.

Taylor,D. L.,Tashakkori,A.,& Hardwick,J. (1996,January). *Teachers' sense of efficacy and attributions for student success*. Paper presented at the annual meeting of the Southwestern Educational Research Association, New Orleans,LA.

Teddlie,C.,Falkowski,C.,Stringfield,S.,Desselle,S.,& Garvue,R. (1984). *The Louisiana School Effectiveness Study: Phase Two*,1982-84. Baton Rouge: Louisiana State Department of Education. (Library of Congress call number: LA, 295,L877,1984)

Teddlie,C.,Kirby,P.,& Stringfield,S. (1989). Effective versus ineffective schools: Observable differences in the classroom. *American Journal of Education*,97(3),221-236.

Teddlie,C.,& Stringfield,S. (1993). *Schools make a difference: Lessons learned from a 10-year study of school effects*. New York: Teachers College Press.

Timberlake,E. M. (1994). Children with no place to call home: Survival in the cars and on the streets. *Child and Adolescent Social Work Journal*,11,259-278.

Trow,M. (1957). Comment on participant observation and interviewing: A comparison. *Human Organization*,16,33-35.

Udry,R. J.,Dole,N.,& Gleiter,K. (1992). Forming reproductive unions in urban Zimbabwe. *International Family Planning Perspectives*,18,10-12.

Ulin,P.,Waszak,C.,& Pfannenschmidt,S. (1996,November). *Integrating qualitative and quantitative research*. Paper presented at Family Health International's Women's Studies Project Technical Advisory Group Annual Meeting, Raleigh,NC.

Virgilio,I.,Teddlie,C.,& Oescher,J. (1991). Variance and context differences in teaching at differentially effective schools. *School Effectiveness and School Improvement*,2(2),152-168.

Wasserman,S.,& Faust,K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Wetzler,S.,Marlowe,D. B.,& Sanderson,W. C. (1994). Assessment of depression using the MMPI, MILLON, and MILLON-II. *Psychological Reports*,75,755-768.

Whyte,W. F. (1943). *Street corner society*. Chicago: University of Chicago Press.

Wittgenstein,L. (1958). *The philosophical investigations*(2nd ed.). New York: Macmillan.

Wolcott,H. F. (1994). *Transforming qualitative data*. Thousand Oaks,CA: Sage.

Yin,R. K. (1993). *Case study research: Design and methods*(2nd ed.). Newbury Park,CA: Sage.

Yin,R. K. (1994). Evaluation: A singular craft. In C. S. Reichardt & S. F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives*(pp. 71-84). San Francisco: Jossey-Bass.

译后记

作为一部方法论的专著，本书特别关注属于当今社会科学发展之前沿的混合方法论，专门讨论了如何将定量研究和定性研究结合起来用于研究实践的问题。

在相对简约的篇幅内，作者首先回顾了社会科学研究领域定性方法与定量方法的历史发展、范式争论和对立主张，进而提出了以实用主义哲学思考为基础的混合方法论理念。在此基础上，作者进一步区分了相对简单的混合方法应用和更高层次的混合模型应用，以相对地位和先后次序为维度，建构了全面的分类体系，以囊括混合方法论的各种实际应用状况。为说明和证明混合方法论的长处和优势，作者又特别根据科学研究的进程结构，逐项介绍了混合方法论在研究设计、抽样和测量计划、资料收集技术，资料分析策略以及成果表达等各个方面的混合方法论问题，并穿插介绍了大量相关的研究案例。由此，作者提出了可实际用于社会科学研究的混合方法论应用指南。

近些年来，定量分析的应用在中国得到了长足的发展，而传统的定性研究也日益走向规范化。然而，两者间源于学术范式和哲学思考的区隔而形成的理念对立，学人中出于学科训练和治学偏好的不同而造就的意见分歧，不仅人为地限定了方法论的阵营界限，形塑了方法论意义上的先设性偏见，而且阻隔了在特定议题上和特定的题域中原本有可能实现的成果交流和优秀成果间的递进式积累。应当承认，这种状况的出现并不利于中国社会科学的发展和进步，也不利于中国社会科学的学者群体对于中国经验和中国模式的认知、理解和解释。在这个意义上讲，本书的翻译引进以及由此所可能引发的思考，无疑可以推进中国学者在超越传统方法论界限的基础上更好地理解方法论的国际前沿发展，更好地应用科学方法论的资源从事对中国现实的学术研究。这不仅是译校者在完成此书译校过程中所体悟的想法，而且是译校者不惮烦难而完成此项译事的信念所在。

其实就本书而论，篇幅之单薄与内容之厚重构成了鲜明的对

比,其文字固简约,其内涵却恢宏精深,书中术语使用之密集,类型建构之全面,是方法论论著中相当少见的。这也是我们在着手本书译事之初时绝未预料到的。更确切地说,我们在本书翻译中所遇到的困难和障碍,所耗费的精力和时间,远远超过了最初的估计。因此,尽管我们付出了努力,并自认已经提供了最佳的译文,但我们仍无法断言其间没有误译和错舛。因而,我们敬请读者不吝赐教。

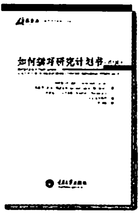
还需说明的是,本书翻译是集体合作的产物。在贾丹硕士提供粗译草稿的基础上,唐海华博士首先对全书进行了严谨通达的翻译;张小劲博士又统校了全书译稿并精心润色改定。全部工作前后费时近一年之久。

当然,本书得以翻译出版,还要感谢重庆大学出版社雷少波先生的努力和敦促!正是雷少波先生和重庆大学出版社在推进万卷方法丛书方面的魄力、热情和投入感动了本书的译校者,促使我们在非常紧张的教学科研之中,仍然非常认真地从事本书的翻译工作。

译校者谨识

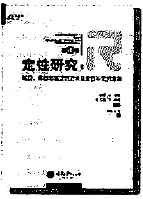
2009 年 6 月 20 日

万卷方法 总策划：雷少波 林佳木



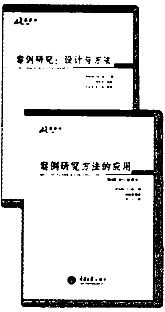
如何撰写研究计划书

研究计划书，也就是我们常说的开题报告。本书的内容分成三个大的部分：1.写作计划书。这部分是本书的核心，不仅提供了适用于所有研究计划书的一般知识，而且还讨论了在研究生教育和资助机构中使用计划书的一些具体问题。2.申请研究资助。3.计划书样例。这部分为我们提供了四个计划书样例，并插入了作者的短评。



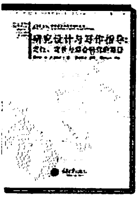
定性研究（全4卷）

本书共4卷，约160万字，内容涉及定性研究的“方法论基础”“策略与艺术”“经验资料收集与分析的方法”“解释、评估与描述的艺术及定性研究的未来”等4个部分。这部在社会科学定性研究方法方面具有里程碑意义的作品，由近80名在国际上享有盛誉的社会科学家，对定性研究的近40个专题进行了讨论，应是所有定性研究方法研习和使用者的案头工具书。



- **案例研究：设计与方法**
- **案例研究方法的应用**

这组有关案例研究方法的经典姊妹篇，已被翻译为多种文字，使国际社会科学界数十万学者受益。
作者罗伯特·K.殷博士，先后取得了麻省和哈佛的学位，并创办了COSMOS调查公司。书中关于案例研究的理论和“案例”，都基于该公司成功实施的美国联邦政府或基金会委托的数百个案例研究项目。
《案例研究：设计与方法》是案例研究的工作流程说明，而《案例研究方法的应用》则展示了在教育、社区研究、法律、戒毒等领域做案例研究的“案例”，为读者提供了可以直接参考的范本。



研究设计与写作指导：定性、定量与混合研究的路径（第2版）

这是一本在西方社会科学界获得广泛赞誉的书，其系统而缜密的研究设计思考、平实易懂的语言、丰富的例子，已经帮助数以十万计社会科学的研究生，顺利完成了他们学位论文的整体设计和写作。

HUNHE FANGFALUN DINGXING FANGFA HE DINGLIANG FANGFA DE JIEHE

- 作为一部方法论的专著，本书专门讨论了如何将定量和定性研究结合起来用于研究实践的问题。

- 在相对简约的篇幅内，作者首先回顾了社会科学研究领域定性方法与定量方法的历史发展、范式争论和对立主张，进而提出了以实用主义哲学思考为基础的混合方法论理念。

- 在此基础上，作者进一步区分了相对简单的混合方法应用和更高层次的混合模型应用，以相对地位和先后次序为维度，建构了全面的分类体系，以囊括混合方法论的各种实际应用状况。

- 为说明和证明混合方法论的长处和优势，作者又特别根据科学研究的进程结构，逐项介绍了混合方法论在研究设计、抽样和测量计划、资料收集技术、资料分析策略以及成果表达等各个方面的混合方法论问题，并穿插介绍了大量相关的研究案例。由此，作者提出了可实际用于社会科学研究的混合方法论应用指南。

发表及参阅相关讨论，请登录：
万卷方法与学术规范博客圈（http//q.blog.sina.com.cn/fafang）

ISBN 978-7-5624-5110-5
9 787562 451105 >
定价：23.00元